Analyzing NFL Managerial Performance Using Sports Data

A Major Qualifying Project
submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirement for the
Degree of Bachelor of Science by

Evan Bettencourt

Thursday, April 25, 2024

Report Submitted to:
Professor Kenny Ching
Professor Randy Paffenroth

Abstract

The importance of advanced analytics and managerial foresight is rapidly growing across professional sports. This project developed insights into player performance in the National Football League (NFL) using advanced player metrics, draft data, and general manager experience from 1970-2023. Using multiple mathematical techniques, including multiple linear regression, fixed effects modeling, and random effects modeling, this study identified critical factors in an NFL general manager's success. The analysis suggests that NFL general managers with more experience in the role are associated with drafting higher-performing players, while those with more executive experience outside of the general manager position are associated with drafting lower-performing players. These findings provide valuable insights for NFL franchises to make informed decisions when employing general managers.

Acknowledgments

I would like to acknowledge Professor Kenny Ching and Professor Randy Paffenroth for their efforts in the completion of this project. My work would not have been possible without the help and resources provided by both professors, and I am thankful for their continuous support throughout the project.

Table of Contents

List of Figures

List of Tables

List of Equations

1.0 Introduction

Great managers are a rare commodity in the global workforce, posing a significant challenge for organizations worldwide. According to Gallup, companies fail to select the correct managerial candidate 82% of the time, highlighting the scarcity of effective leadership (Harter & Beck, 2023). Weak management is particularly concerning because it is correlated with decreased employee productivity, engagement, and commitment.

This project focuses on managerial performance in the National Football League (NFL). As the most popular North American professional sports league, the pressure for each NFL team to succeed has never been greater. Consequently, team owners want capable managers who can lead effectively and bring long-term success to their franchises. This project aims to provide insights into the role of National Football League general managers in identifying successful long-term player performance.

2.0 Background

In 2023, over 150 million people in the United States watched live sports content at least once per month (Stoll, 2023). This number showcases the popularity and cultural significance of sports within the United States, as nearly half of the country's population regularly tunes in to watch. The most popular sports are the four major North American leagues: the National Football League (NFL), National Basketball Association (NBA), National Hockey League (NHL), and Major League Baseball (MLB). Across these leagues, over $45 million in revenue was created in 2023 (A, 2024). The NFL led with $18.6 million in revenue, followed by the NBA, MLB, and NHL with $10.58 billion, $10.32 billion, and $5.93 billion, respectively. These revenues come from various sources, including television contracts, brand sponsorships, licensing deals, concessions, and merchandise. This project focuses on analyzing aspects of the National Football League, so it will be discussed in further detail.

2.1 The National Football League

The National Football League (NFL) is a major professional gridiron football league in the United States that was founded in 1920 in Canton, Ohio. The league began with just 14 teams in its inaugural season and has since grown to 32 teams across the country. Only two of the original teams remain in today's NFL, reflecting years of instability and competition from rival organizations.

2.1.1 League Organization History

As the NFL began to gain popularity, the American Football League (AFL) sought to rival it as the largest and most popular gridiron football league. The AFL began play in 1960, introducing many new features that the NFL did not have, such as official time on the scoreboard clock, optional two-point conversions after touchdowns, and names on player jerseys (Frommer, 2023). The two leagues engaged in bidding wars for the best players, creating uncertainty in the drafts of both leagues. In 1966, The NFL and AFL agreed on a merger to combine the two leagues, their drafts, and their championship games. This became official in 1970, when the NFL and AFL merged into the modern-day National Football League, centralizing all the best football talent in the world into one league.

At the time of the NFL/AFL merger, the new league began with 26 teams split into two conferences - the American Football Conference (AFC) and the National Football Conference (NFC). Each conference had three divisions with an unequal distribution of teams. As the NFL grew, the league added expansion teams and balanced the divisions. In 1976, the league expanded to 28 teams with the addition of the Seattle Seahawks and Tampa Bay Buccaneers. In 1995, the Carolina Panthers and Jacksonville Jaguars were added, bringing the total to 30 teams with balanced five-team divisions. However, in 1999, the Cleveland Browns franchise was reinstated into the AFC Central, bringing the league up to 31 teams and causing imbalance once again.

In addition to league expansion and organization, relocation in the NFL has been a significant part of its history. Since the NFL's founding in 1920, 11 franchises have relocated, the most recent being the Raiders' move from Oakland to Las Vegas in 2020 (Kerr, 2020). Franchises relocate for various reasons, including seeking better stadium deals, reaching larger markets, or improving the team's financial situation.

After decades of imbalance and imperfection, the 2000s marked a turning point for the NFL. In 2002, the Houston Texans were added as an expansion team, and there have been no more additions to the league since. With 32 total teams, the entire layout of the NFL was reimagined. In the modern-day NFL, the 32 teams are split equally into the AFC and NFC. Each conference has 16 teams divided equally into four divisions: East, North, South, and West, with four teams in each division. This division structure has proven successful for over 20 years as the NFL has experienced exponential growth. Despite rumors of further expansion to continue this growth, the idea is often dismissed due to the current success of the 32-team system.

2.1.2 Viewership, Revenue, and the Growing Drive to Succeed

The NFL is not only the most popular professional sports league for television viewers but also some of the most popular television overall. Over the past four years, at least 70 of the 100 most-watched United States TV broadcasts have been NFL games (Figure 1). This staggering number showcases the true dominance of the league and its ability to captivate viewers.

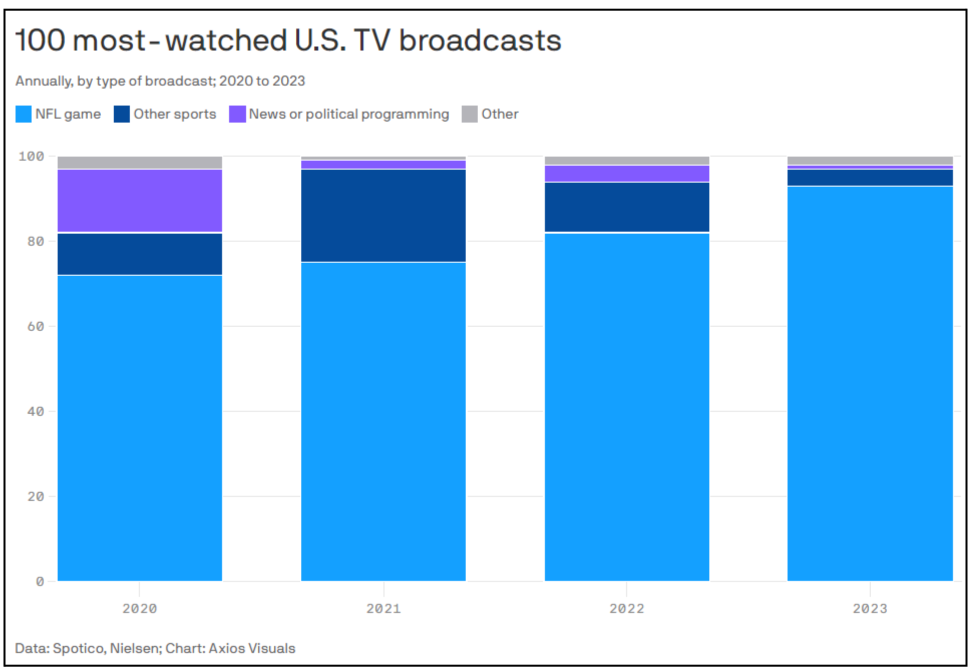

Figure 1: 100 Most-watched U.S. TV Broadcasts 2020-2023 (Axios Visuals, 2024)

Because the NFL only has 17 games for each team in the regular season, each game is vital to a team's success. This added importance to each game creates more excitement for fans to watch every week. Over the past 15 years, total viewership has gone up and down but has seen tremendous growth as of late. In 2023, the average NFL regular season game had 17.9 million viewers (Figure 2).

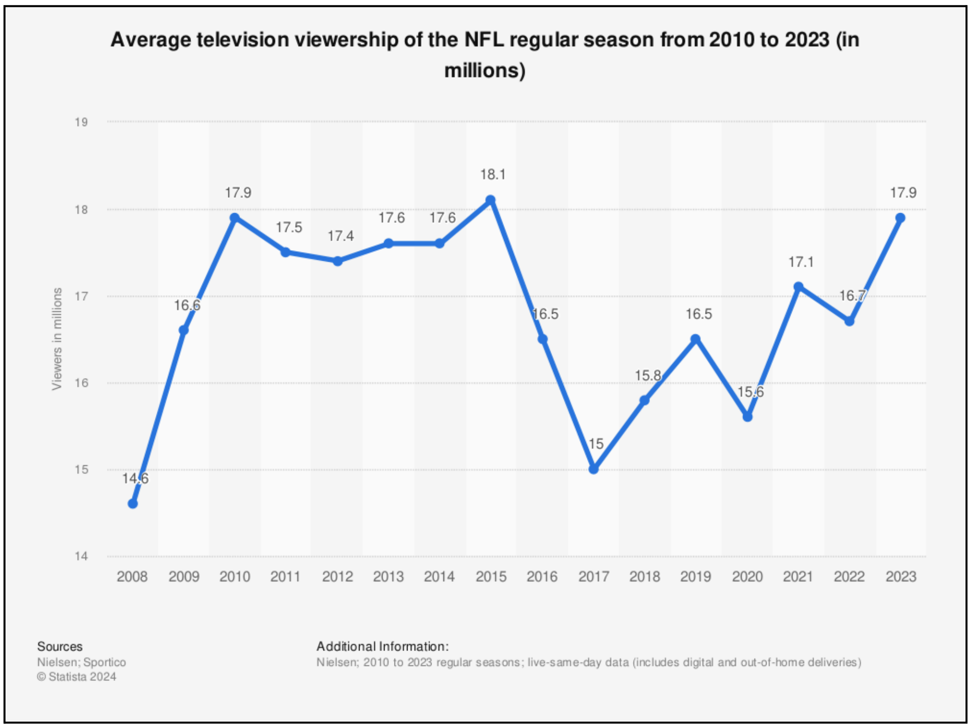

Figure 2: Average Television Viewership of the NFL Regular Season from 2010 to 2023 (Sportico, 2024)

As NFL games continue to gain popularity, the league makes sure to capitalize on revenue opportunities. Television contracts are the largest source of revenue for the league, with the current contract bringing in more than $12 billion per year from 2023-2033 (Ozanian, 2021). Every NFL team is looking to benefit from growing viewership and revenue across the league. To do so, teams must focus on being successful. According to a study from Claremont McKenna

College (Pautler, 2010), a higher standard deviation of winning percentage over ten years was associated with lower revenues. Over the long term, fans want to see teams consistently winning and are willing to pay larger amounts of money to watch winning teams. Because of these factors, the desire for teams to succeed has never been higher. One of the most important aspects of a successful team is having a strong general manager to lead it. For general managers to succeed and avoid being replaced, they must be able to build a winning team through strong player evaluation, most importantly through the NFL draft.

2.1.3 The NFL Draft

The NFL draft is an annual event that allows teams to acquire new talent from the National Collegiate Athletic Association (NCAA). The most important aspect of the NFL draft is how the draft picks are assigned. In the current format, teams can receive two types of draft picks. First are the “regular” draft picks where each of the 32 teams receives one pick in each of the seven rounds. The NFL systematically distributes the draft selections to each team based on their performances in the previous season. First, teams that did not reach the playoffs are assigned picks 1-20 using the reverse order of their final regular season records. If multiple teams end the regular season with the same record, the determination of draft position is decided by the strength of schedule — the aggregate winning percentage of a team’s opponents (National Football League, n.d.-b). The team that played the schedule with the lower winning percentage is awarded the higher pick in the tiebreaker. If teams have the same strength of schedule, the NFL uses division and conference tiebreakers. This list of tiebreakers is long and comprehensive to ensure that all possible scenarios are covered.

Next, teams that did reach the playoffs are assigned picks 21-32 based on playoff performance. The rules for draft pick determination can be found in Figure 3.

- The four teams eliminated in the wild card round pick in slots 21-24 in the reverse order of their final regular season records.
- The four teams eliminated in the divisional round pick in slots 25-28 in the reverse order of their final regular season records.
- The two teams that lost in the conference championships pick in the 29th and 30th spots in the reverse order of their final regular season records.
- The team that lost the Super Bowl has the 31st pick in the draft.
- The Super Bowl champion has the 32nd and final spot in each round.

Figure 3: Draft Pick Determination for Playoff Teams (National Football League, n.d.-b)

In addition to the “regular” draft picks, the NFL also has a system of “compensatory free agent” picks to help clubs that have lost free agents to another team. In any given year, the NFL can assign up to 32 of these additional picks that can take place at the end of rounds 3-7. These selections are determined by an algorithm that considers player salary, playing time, and postseason honors (National Football League, n.d.-b). The compensatory pick system is a measure the league has taken to support competitive equity and is critical to the draft format.

Once NFL teams are assigned their selections, they become valuable assets for the team’s future. Draft picks can be used in two key areas: selecting draft-eligible players or trading to other clubs. To be draft-eligible, a player must have been out of high school for at least three years and have used up their college eligibility before the start of the next NCAA season, or get approval to enter the draft early from the league (National Football League, n.d.-b). Draft selections are a currency for NFL teams, and every pick has a different value. As the key asset for acquiring new talent on an NFL team, draft selections are a critical focus as the desire to succeed grows.

2.2 The Rise of Sports Analytics

In the modern professional sports landscape, most teams across all major leagues use analytics to gain a competitive advantage. However, it was not always this way. The utilization of modern sports analytics developed in 1974 when Bill James, Dick Cramer, and Pete Palmer co-founded the Society of American Baseball Research's (SABR) Statistical Analysis Committee to analyze the performance of Major League Baseball teams (Mizels et al., 2022). Initially, the idea of using analytics in sports was rejected for interrupting traditional practices. However, sports analytics would see strong success and rapidly grow as a field.

The first breakthrough success of sports analytics was in the late 1990s and early 2000s by Billy Beane and the "Moneyball" Oakland Athletics. As the general manager of an MLB team in a small market with a low budget to spend on player salaries, Beane had to find unique ways to build a successful team. Using the Pythagorean winning percentage equation developed by Bill James, Beane and his team were able to estimate the expected winning percentage of his team using two simple statistics, runs scored and runs allowed. This equation can be seen in Equation 1 below.

$$Expected\ Win\ Percentage\ = \frac{(Runs\ scored)^2}{(Runs\ scored)^2 + (Runs\ allowed)^2}$$

Equation 1: Pythagorean Winning Percentage (Major League Baseball, n.d.)

This formula allowed Beane to project how many runs his team needed to score and how many runs his team could allow to reach a certain expected winning percentage. Based on historical data, he could estimate what winning percentage the team would need to make the playoffs and base his projections on that expectation. Using this knowledge and more advanced metrics created by Bill James, Beane broke down what statistics correlated with runs scored and runs allowed. Beane then focused on acquiring players who were overlooked and undervalued by other teams because of their age, injury history, and looks. This team-building strategy brought the Oakland Athletics from a bottom-end MLB team in the 1990s to a perennial contender in the early 2000s. Beane's success showcased the power of sports analytics to those around him, shifting the landscape of sports strategy forever. Teams began to invest heavily in analytics departments, bringing in statisticians, data scientists, and computer programmers to help make data-driven decisions based on the success of the Oakland Athletics and Billy Beane.

In addition to individual teams, sports leagues are investing in analytics. In the NFL, the league has an analytics team, NFL Next Gen Stats. Next Gen Stats partners with Amazon Web Services, Zebra Technologies, and Wilson Sporting Goods. They do so with a tracking system in every NFL stadium consisting of ultra-wideband receivers, RFID tags installed in players' shoulder pads, and RFID tags on officials, pylons, sticks, chains, and the ball (National Football League, n.d.-a). With all of this technology, the NFL can provide teams with a variety of advanced metrics to use as they please.

As the landscape of professional sports continues to implement analytics in more areas, it is more important than ever that all members of an organization understand how to take advantage of it. Data-driven decision-making is gaining traction in a variety of industries, and the NFL is no different. General managers need to understand how they can use sports analytics to improve their team's chances of winning and gain an edge over the rest of the league.

2.3 Managerial Foresight

Managerial foresight refers to predicting how a manager's actions can create a competitive advantage for an organization (Amsteus, 2008). According to a study on the goals of sporting club ownership, success is defined using the "eight Ps" of performance, profit, platform, preemptive, purpose, profile, power, and passion (Foster et al., 2023). While there are many factors in this list, the authors note that performance is often a primary motive. With winning at the top of the priority list for NFL owners, teams are looking for advantages to become more successful, and strong leaders are critical. The two main leaders of an NFL team are the head coach and the general manager. The head coach is responsible for coaching the players on the team and getting the best performance out of them. The general manager is responsible for roster management, which includes the NFL draft, free agency, salary cap management, and trades. This analysis will focus on one aspect of roster management, the NFL draft, and what general manager characteristics may lead to more successful player selections.

3.0 Design Process

Before analyzing the performance of NFL general managers, data were collected and cleaned through a series of steps to create a comprehensive dataset.

3.1 Data Collection

To create a dataset with all the information necessary for analysis, data were collected for the NFL draft, NFL player performance, and NFL general manager characteristics from 1970-2023. Databases from Pro Football Reference, Pro Football Focus (PFF), and RetroSeasons were used. Pro Football Reference is an analytics website that contains complete player, team, and league stats dating back to 1920, the original founding of the NFL. Pro Football Focus is a sports analytics company focused on thorough and unique player grades and rankings for the NFL. RetroSeasons is a website that recaps historical sports seasons for every professional sports league. Each site made a critical contribution to creating a dataset for this project.

For NFL draft data, Pro Football Reference draft datasets were used. The important fields for this analysis included player name, a unique player identifier, draft round, overall draft selection, team, and position. Each draft year's data from 1970-2023 was exported as .csv files and combined into one table in Microsoft Excel. Pro Football Reference's yearly statistic datasets were used for player performance data. This included passing, rushing, receiving, and defense statistics from 1970-2023. There were hundreds of datasets compiled into one square dataset to track player performance. However, this dataset was missing information for offensive linemen, creating the need for a new player performance evaluation method.

To allow for all offensive and defensive players to be included in the player performance analysis, Pro Football Focus's season-level player grades were used. These player grades have become a widely accepted metric across the NFL since beginning in 2006 (Pro Football Focus, n.d.-b). The PFF grading system is designed to grade every player on every play during an NFL game, providing extreme attention to detail that cannot be found in typical counting statistics such as completions, interceptions, or sacks. The goal of the grading system is to grade the play, not the result of the play. For example, if a quarterback makes a great throw that is dropped by his wide receiver, traditional statistics would not give the quarterback any credit. However, PFF's grading system would consider this a positive play for the quarterback since he threw the ball well to the correct spot. The benefit of this system is that it allows for a larger sample size of data to tell the full story of a player's performance that may not be fully captured by traditional statistics (Pro Football Focus, n.d.-b).

To grade each player on a given play, a scale from -2 to +2 in 0.5 increments is used with a grade of 0 considered as the "expected" grade for a play. PFF collects over 200 fields of data on every play of every game, allowing for a strong estimate to be created for the expected outcome of each play, the baseline for the grading system. Figure 4 shows an example of the scale, with a game-ending interception from a quarterback as a -2 grade, a perfectly thrown deep pass into tight coverage on the game-winning drive as a +2 grade, and other positively and negatively rated plays in between.

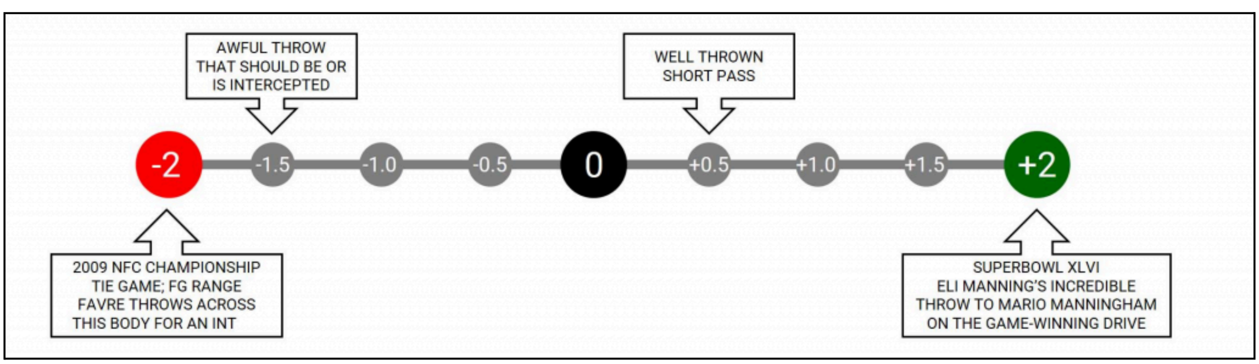

Figure 4: PFF Grading Scale Example (Pro Football Focus, n.d.-a)

At the end of each game and each season, the plus-minus grades from the previously mentioned system are converted to a 0-100 scale. The season grades are not simply the average of the grades a player receives in a season, they consider the player's performance over the entire season. For example, there have been hundreds of games where a quarterback has thrown for three touchdowns, but if a quarterback threw for three touchdowns in every game in a season, it would be considered one of the best seasons of all time. This final season grade is the metric used for player performance analysis. Total offense PFF grade was used for offensive players, while total defense PFF grade was used for defensive players. While PFF's grading system is subjective and only tracked since 2006, it is one of the most detailed continuous variables available to evaluate player performance in the NFL.

Finally, data on general managers from 1970-2023 and their collegiate and pro football experiences were collected from RetroSeasons and Pro Football Reference to determine the independent variables for analyzing managerial performance. From these websites, information was collected on NFL general manager experience, prior NFL playing experience, NFL executive experience, NFL head coaching experience, and NCAA head coaching experience. The next step of the design process was to clean this data to allow for successful joins before analyzing.

3.2 Data Cleaning

In the data collection process, multiple data cleaning steps were completed to connect all data into one usable dataset. First, abbreviations for teams that have relocated since 1970 are different for each location they played in, even though the team is considered the same organization. To create consistency across locations for the same organization, all team abbreviations were updated to abbreviations for the most recent season, 2023. Additionally, Pro

Football Reference and PFF use different abbreviations for many of the same organizations. Thus, all abbreviations were updated to match Pro Football Reference's three-letter style for simplicity in the dataset.

The second key area for data cleaning was player names. In the Pro Football Reference draft dataset and PFF player performance dataset, differences included hyphenated names, preferred names, and suffixes. To fix this error, the Pro Football Reference and PFF player performance data were connected by name, year, and team. All players that had data errors occur during this match had their names examined between the two datasets. For every instance of a name that failed to match, the datasets were cross-examined to find the mismatched names from each, and the name from the PFF dataset was used as the final value. After updating the names of all players, data errors were eliminated and all data were ready to be joined together.

Multiple steps were taken to combine the draft, player performance, and general manager data into one dataset. First, the draft data were joined with the general manager data using the year and team of the draft and matching it with the year and team of the general manager. Any players without a match were players who never got drafted. These players were removed from the dataset since the goal is to examine general manager draft performance. Next, the data were joined with the Pro Football Reference player performance data using the unique player code. Finally, the PFF player grades were joined onto the dataset using player name, team, year, and position, as the unique player codes from the Pro Football Reference and PFF sites were inconsistent. While this method of connecting data is not perfect, as multiple players could have the same values in all four categories, it did work for this sample as there were zero occurrences of that scenario. To create a more accurate system, a truly unique player identifier is necessary, but the two datasets used different identifiers. Due to the time restrictions on this project, creating my own unique identifier would not have been viable due to the number of records that would have to be scanned for the same name and manually changed.

After completing all data cleaning and joining, the final dataset included over 20,000 individual player seasons, 7,200 NFL players, and 150 NFL general managers. Table 1 lists each variable in the final dataset and describes each.

Table 1: Final Dataset Data Dictionary

Variable	**Description**
player_name	Player name
player_code	Unique player identification code
year	Year of the season kickoff
team	Team abbreviation
year_team	Concatenated year and team
name_year_team	Concatenated player name, year, and team
age	Player's age on December 31st of the given year
pos	Position
pff_grade	Pro Football Focus player grade
draft_gm_name	Name of the general manager who drafted the player
draft_gm_code	Unique identifier of the general manager who drafted the player
pick	Overall draft selection of the player
gm_exp	General manager experience of the general manager who drafted the player
nfl_exec_exp	NFL executive experience (outside of the general manager role) of the general manager who drafted the player (e.g. scout, player personnel, pro personnel, vice president)
nfl_pl_exp	NFL player experience of the general manager who drafted the player

nfl_hc_exp	NFL head coach experience of the general manager who drafted the player
ncaa_hc_exp	NCAA head coach experience of the general manager who drafted the player

4.0 Data Analysis Methodology

In this data analysis, it is assumed that a "better" player is a player with a higher PFF grade. While the metric is not perfect, PFF grade is widely accepted as a strong metric to compare all NFL players regardless of position. There are very few metrics available to compare NFL players across positions due to the drastic differences in the roles and responsibilities of each, but the PFF grade can.

Throughout the analysis, assumptions had to be made to assess all general managers equally. Some scenarios cannot be accounted for using only quantitative results data, including the degree of control awarded to a general manager in the draft, player injuries, and player off-the-field issues. This analysis assumes that each general manager is fully responsible for the results of each draft. Additionally, it is assumed that player injuries and off-the-field issues are uncontrollable by a team's general manager. With these assumptions, the performance of each NFL general manager can be analyzed equally.

4.1 Multiple Linear Regression

The first method used to analyze managerial performance was multiple linear regression. Multiple linear regression attempts to model the relationship between two or more independent variables and a single continuous dependent variable by fitting a linear equation to the data (Hanck, 2024). Equation 2 represents the model for multiple linear regression.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i$$

Equation 2: Multiple Linear Regression (Hanck, 2024)

Where:

- Y_i is the dependent variable Y for the ith observation

- β_0 is the intercept term
- $X_{1i}, X_{2i}, \dots, X_{ki}$ are k distinct independent variables for the ith observation
- $\beta_1, \beta_2, \dots, \beta_k$ are the estimated regression coefficients
- u_i is the error term

Ordinary Least Squares (OLS) regression was used to test the multiple linear regression model due to its availability and popularity in Python code. This model is a method of linear regression that selects values for the least squares estimator β such that the sum of the squared residuals ($\sum_{i=1}^{n} (u_i)^2$ or $u'u$) is minimized.

The derivation of the least squares estimator β used in this regression analysis can be found in *Appendix A*. To implement the multiple linear regression model in practice, the OLS model in the statsmodels Python library was used (see *Appendix B* for code).

In the first attempt using this model, all potential explanatory variables were tested. This included general manager experience (*gm_exp*), NFL player experience (*nfl_pl_exp*), NFL executive experience (*nfl_exec_exp*), NFL head coach experience (*nfl_hc_exp*), and NCAA head coach experience (*ncaa_hc_exp*). Additionally, overall draft selection (*pick*) was included in all analyses to control for the higher expected value of a higher-drafted player. Following the first iteration of this analysis, all statistically insignificant variables were removed except NFL player experience. This variable is included in all future analyses to compare the results of this project with a previous iteration that analyzed managerial performance in the NBA (Bissonette, 2022). Using multiple linear regression is a strong start for analyzing the importance of various general manager characteristics, but other methods can be used to improve the effectiveness of the analysis.

4.2 Fixed Effects Regression

The next data analysis method used in this project was fixed effects regression. This regression style adjusts for heterogeneity among entities and time and avoids omitted variable bias (Hanck, 2024). In this analysis, the entities are the individual players and the time is the year the given season began. This regression method was chosen to test if fixed effects influence player performance in the NFL. An example of an entity fixed effect in the NFL is an individual's skill and an example of a time fixed effect is a change in the league rules. It is important to account for entity and time fixed effects to eliminate bias they may have on player performance.

The regression model for fixed effects regression is similar to the model for multiple linear regression in Equation 2 with some changes. A fixed effects model evaluates each independent variable for a given entity i at a certain time t. Additionally, group-level dummy variables are introduced to control for categorical variables that cannot be directly included as independent variables in the model. These dummy variables are treated as independent variables in the regression. For a regression analysis with T entities or time periods, $T - 1$ dummy variables are used to ensure that the model captures the group effects but does not lead to perfect multicollinearity among variables. Equation 3 represents the entity and time fixed effects model.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \ldots + \beta_k X_{kit} + \gamma_2 D2_i + \ldots + \gamma_T DT_i + \delta_2 B2_t + \ldots + \delta_T BT_t + u_{it}$$

Equation 3: Time and Entity Fixed Effects Regression (Hanck, 2024)

Where:

- Y_{it} is the dependent variable Y for entity i at time t
- β_0 is the intercept term

- $\beta_1, \beta_2, \ldots, \beta_k$ are the estimated regression coefficients
- $X_{1it}, X_{2it}, \ldots, X_{kit}$ are the k independent variables for entity i at time t
- $D2_i, \ldots, DT_i$ are dummy variables representing entity fixed effects. $D2_i$ represents entity i being in group 2, following the same pattern up to DT_i representing entity i in group T
- $\gamma_2, \ldots, \gamma_T$ are the coefficients for the entity fixed effects
- $B2_t, \ldots, BT_t$ are dummy variables representing time fixed effects. $B2_t$ represents time t being in period 2, following the same pattern up to BT_t representing time t in period T
- $\delta_2, \ldots, \delta_T$ are the coefficients for the time fixed effects
- u_{it} is the error term

While this model encapsulates entity and time fixed effects, we run into an error when implementing it in Python. In this model, including entity fixed effects is not feasible because there is no variation among independent variables (general manager experiences) for each entity. Each player is drafted by one general manager whose experience as a general manager, prior NFL playing experience, and NFL executive experience are time-invariant. While a general manager can gain more experience over time, the instance of their experience at the time of the draft will remain constant throughout the given drafted player's career. Thus, we must remove the entity fixed effects and use a fixed effects equation that only includes time fixed effects. Equation 4 represents the time fixed effects model.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \ldots + \beta_k X_{kit} + \delta_2 B2_t + \ldots + \delta_T BT_t + u_{it}$$

Equation 4: Time Fixed Effects Regression (Hanck, 2024)

Where:

- Y_{it} is the dependent variable Y for entity i at time t

- β_0 is the intercept term
- $X_{1it}, X_{2it}, \dots, X_{kit}$ are the k independent variables for entity i at time t
- $\beta_1, \beta_2, \dots, \beta_k$ are the estimated regression coefficients
- $B2_t, \dots, BT_t$ are dummy variables representing time fixed effects. $B2_t$ represents time t being in period 2, which follows the same pattern up to BT_t representing time t in period T
- $\delta_2, \dots, \delta_T$ are the coefficients for the time fixed effects
- u_{it} is the error term

To run this model only with time fixed effects to control for variables that are constant across entities but vary over time, the PanelOLS model from the statsmodels Python library was used with time effects set to "true" and entity effects set to "false" (see *Appendix C* for code).

While fixed effects modeling helps account for time-specific factors that impact all NFL players in the same way at a given time, it cannot do the same for entity-specific factors due to the nature of the dataset. Because of this, a different approach needs to be taken to help account for potential entity-specific effects.

4.3 Mixed Effects Regression

Due to the time-invariant nature of the managerial characteristics with respect to the players, random effects modeling is appropriate for the entity-specific effects in this analysis (The Trustees of Princeton University, n.d.). To combine time fixed effects from the previous regression with entity random effects, a mixed effects regression model was used. This model eliminates bias from unobserved influences that change over time but are constant over entities and controls for factors that differ across entities but are constant over time (Hanck, 2024).

The mixed effects model is an extension of the multiple linear regression model from Equation 2, adding fixed effects and random effects for a given clustering variable. In this analysis, the clustering variable is *player_code*, which groups records of player performance for every season an individual player played. Equation 5 represents the linear mixed effects regression model.

$$Y = X\beta + Z\gamma + \epsilon$$

Equation 5: Linear Mixed Effects Regression ("Mixed Effect Regression", n.d.)

Where:

- Y is the dependent variable
- X is a $\mathbb{R}^{N\times p}$ matrix containing the observed values for each individual (N) for each independent variable (p)
- β is a $\mathbb{R}^{p\times 1}$ vector containing the regression coefficients for each independent variable (p)
- Z is a $\mathbb{R}^{N\times q}$ matrix containing the observed values for each individual (N) for each covariate (q)
- γ is a $\mathbb{R}^{q\times 1}$ vector containing the regression coefficients for each covariate (q)
- ϵ is the residual error

To analyze the dataset with this model in Python, the mixedlm model from the statsmodels library was used. In this model, the regression formula was specified with *pff_grade* as the dependent variable and *pick*, *gm_exp*, *nfl_exec_exp*, *nfl_pl_exp*, and *year* as independent variables. The variable *year* is defined as a categorical variable to allow the model to estimate a separate effect for each year. The Python code for the mixed effects regression can be seen in *Appendix D*.

5.0 Results

This section details the results of each regression test conducted. The testing was an iterative process, continually revising methodology and determining the best technique to analyze the effects of managerial characteristics on NFL player performance.

5.1 Multiple Linear Regression

Using the OLS linear regression model in Python with the seven potential explanatory variables along with the overall draft selection number to control for higher drafted players' expectation to perform better, we obtain the results in Table 2 below.

Table 2: Multiple Linear Regression First Iteration Results

Variable	**Coefficient**	**Standard Error**	**P-value**
pick	-0.0314	0.001	< 0.001
gm_exp	0.0806	0.012	< 0.001
nfl_exec_exp	-0.0296	0.010	0.003
nfl_pl_exp	-0.0925	0.227	0.684
nfl_hc_exp	0.0155	0.020	0.427
ncaa_hc_exp	-0.0276	0.047	0.557

The results in Table 2 show that a few potential explanatory variables have no statistical significance due to their P-values being larger than 0.25. These variables include NFL player experience (*nfl_pl_exp*), NFL head coach experience (*nfl_hc_exp*), and NCAA head coach experience (*NCAA_hc_exp*). With these results from the first regression, we move forward to the second iteration of testing where all statistically insignificant variables were removed except NFL player experience (*nfl_pl_exp*) to compare the results of this project with a previous iteration of this project that analyzed managerial performance in the NBA (Bissonette, 2022). The results for this iteration can be seen in Table 3.

Table 3: Multiple Linear Regression Second Iteration Results

Variable	**Coefficient**	**Standard Error**	**P-value**
pick	-0.0314	0.001	< 0.001
gm_exp	0.0852	0.010	< 0.001
nfl_exec_exp	-0.0329	0.008	< 0.001
nfl_pl_exp	-0.1104	0.223	0.621

In this iteration of regression testing, all variables proven to be statistically significant in the first iteration of this regression have a P-value of 0, showing that the test results are statistically significant. There are minute differences between these results and the previous iteration. Here, we can see that the coefficient for *gm_exp* increased slightly and the coefficient for *nfl_exec_exp* decreased slightly. These changes further support the conclusions made from the first iteration of regression testing.

5.2 Fixed Effects Regression

With the time fixed effects PanelOLS model in Python, the following results for independent variables are obtained (Table 4).

Table 4: Time Fixed Effects Regression Results

Variable	Coefficient	Standard Error	P-value
pick	-0.0313	0.0019	< 0.001
gm_exp	0.0846	0.0166	< 0.001
nfl_exec_exp	-0.0339	0.0141	0.0195
nfl_pl_exp	-0.1468	0.3687	0.6905

From this analysis, it can be seen that the coefficient values for the independent variables are slightly different than the results from the multiple linear regression analysis. The negative effect of overall draft selection (*pick*) and PFF grade as well as the positive effect of general manager experience (*gm_exp*) and PFF grade decreased in magnitude after including time fixed effects. However, the negative effect of NFL executive experience (*nfl_exec_exp*) and PFF grade increased in magnitude after including these effects. Alongside this change in magnitude, the P-value for NFL executive experience increased to not be statistically significant at the 1% level, but still significant at the 5% level. Additionally, the effect of NFL player experience (*nfl_pl_exp*) remains statistically insignificant after including time fixed effects due to its high P-value. These results show that time fixed effects in the NFL have a small impact on the expected performance of players, but are not a major contributor.

5.3 Mixed Effects Regression

Using the mixed effects regression model in Python, the following results for independent variables and intercept were obtained (Table 5).

Table 5: Linear Mixed Effects Regression Results

Variable	**Coefficient**	**Standard Error**	**P-value**
pick	-0.030	0.002	< 0.001
gm_exp	0.067	0.015	< 0.001
nfl_exec_exp	-0.025	0.013	0.053
nfl_pl_exp	-0.300	0.334	0.369

After including both time fixed effects and entity random effects using the mixed effects regression model, the effects of all independent variables decreased in magnitude from the previous regression testing. In this final test, the P-value for NFL executive experience increased to no longer be significant at the 5% level but still is significant at the 10% level. The results also show that NFL player experience remains statistically insignificant after including both entity-specific and time-specific effects in the regression model. Now that these effects have been considered in the analysis, we can be more confident with the obtained results and can make three definitive conclusions about managerial success in the NFL:

1. General managers with more experience as a general manager are associated with drafting higher-performing players
2. General managers with more experience as an NFL executive outside of the general manager position are associated with drafting lower-performing players
3. Prior experience as an NFL player, NFL head coach, or NCAA head coach has no statistically significant impact on the association between a general manager and the performance of the players they draft

6.0 Conclusion

From the obtained results of the mathematical analysis, three conclusions were made. The first conclusion discovered was that general managers with more experience as a general manager are associated with drafting higher-performing players. This conclusion seems intuitive, as general managers with more experience have participated in many drafts and evaluated more players. Because of this, they may be more likely to recognize trends in the NFL draft and capitalize on their additional knowledge, leading to higher performance compared to other managers around the league.

The second conclusion found in this project was that general managers with more experience as NFL executives outside of the general manager position are associated with drafting lower-performing players. This conclusion is interesting to discover since general managers with more executive experience have a longer track record of valuable NFL experience before becoming a general manager. However, these managers may have progressed through the executive ladder slower for a reason, which could explain their poorer performance as general managers when they eventually reach that rank.

The final conclusion discovered in this analysis is that a general manager's previous experience as an NFL player, NFL head coach, or NCAA head coach is statistically insignificant to the player performance of the players they draft. The insignificance of prior NFL player experience is counterintuitive to a similar analysis in the NBA, which found that prior player experience was associated with drafting higher-performing players (Bissonette, 2022). However, the insignificance of head coaching experience at both the NFL and NCAA levels is understandable, as coaching and managing require different skill sets to be successful.

To further investigate managerial performance in the National Football League, one could analyze the salary cap management of general managers. The salary cap is a rule that restricts the amount of money a team can spend on player contracts, which is designed to maintain a competitive balance in the league. General managers must decide how to distribute contracts to players to improve their team's performance while staying under the salary cap. By comparing player contracts with player performance, it could be determined if general managers are overpaying, underpaying, or fairly paying for players. This research could complement the results from this project and provide more conclusive results on overall managerial performance in the NFL.

7.0 References

A, M. (2024, January 9). *Richest Sports Leagues in the World - Top 10 Fanatical Following*. Fresherslive. https://www.fresherslive.com/latest/articles/richest-sports-leagues-in-the-world-top-10-fanatical-following-1555237379

Amsteus, M. (2008, February 22). *Managerial foresight: Concept and measurement*. Foresight. https://www.emerald.com/insight/content/doi/10.1108/14636680810856026/full/html

AV. (2015, February 15). *Derivation of the least squares estimator for beta in matrix notation*. Economic Theory Blog. https://economictheoryblog.com/2015/02/19/ols_estimator/

AV. (2018, October 17). *Derivation of the least squares estimator for beta in matrix notation – proof nr. 1*. Economic Theory Blog. https://economictheoryblog.com/2018/10/17/derivation-of-the-least-squares-estimator-for-beta-in-matrix-notation-proof-nr-1/

Axios Visuals. (n.d.). *100 most-watched U.S. TV broadcasts*. Axios.com. Retrieved from https://www.axios.com/2024/01/07/nfl-tv-ratings-live-events-viewership.

Bissonette, J., & Francis, C. (2022). Studying Managerial Foresight in Sports Analytics. : Worcester Polytechnic Institute.

Foster, G., O'Reily, N., Devereux, J. B., & Shundi, M. (2023, November 6). *General manager and head coach exits in the NBA and the NFL*. Sport, Business and Management: An

International Journal. https://www.emerald.com/insight/content/doi/10.1108/SBM-09-2022-0081/full/html

Frommer, F. (2023, July 21). *American Football League*. Encyclopædia Britannica. https://www.britannica.com/topic/American-Football-League-American-gridiron-football-organization

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2024, February 13). *Introduction to Econometrics with R*. Econometrics with R. https://www.econometrics-with-r.org/index.html

Harter, J., & Beck, R. J. (2023, July 21). *Why great managers are so rare*. Gallup.com. https://www.gallup.com/workplace/231593/why-great-managers-rare.aspx

Kerr, J. (2020, February 4). *Las Vegas Raiders latest NFL franchise to change cities: Here's a look at all 11 who've done it*. CBS Sports. https://www.cbssports.com/nfl/news/las-vegas-raiders-latest-nfl-franchise-to-change-cities-heres-a-look-at-all-11-whove-done-it/

Major League Baseball. (n.d.). *Pythagorean winning percentage: Glossary*. MLB.com. https://www.mlb.com/glossary/advanced-stats/pythagorean-winning-percentage

Mixed Effect Regression. Python for Data Science. (n.d.). https://www.pythonfordatascience.org/mixed-effects-regression-python/

Mizels, J., Erickson, B., & Chalmers, P. (2022, August). *Current state of data and analytics research in baseball*. Current reviews in musculoskeletal medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9276858/

National Football League. (n.d.-a). *NFL next gen stats*. NFL Football Operations. https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/

National Football League. (n.d.-b). *The Rules of the NFL Draft*. NFL Football Operations. https://operations.nfl.com/journey-to-the-nfl/the-nfl-draft/the-rules-of-the-draft/

Ozanian, M. (2023, August 30). *Why the NFL could reap more than $126 billion in TV money by 2033*. Forbes. https://www.forbes.com/sites/mikeozanian/2023/08/30/why-the-nfl-could-reap-more-than-126-billion-in-tv-money-by-2033/?sh=4b2d685215b5

Pautler, M. D. (2010). *The Relationship Between Competitive Balance and Revenue in America's Two Largest Sports Leagues* (thesis).

Pro Football Focus (n.d.-a). *PFF Grading Scale Example*. (n.d.). PFF. Retrieved from https://www.pff.com/grades

Pro Football Focus. (n.d.-b). *PFF player grades*. PFF. https://www.pff.com/grades#:~:text=The%20grades%20allow%20for%20easy,is%20crucial%20when%20evaluating%20players

Sportico. (January 10, 2024). Average television viewership of the NFL regular season from 2010 to 2023 (in millions) [Graph]. In *Statista*. Retrieved April 04, 2024, from https://www-statista-com.ezpv7-web-p-u01.wpi.edu/statistics/289979/nfl-number-of-tv-viewers-usa/

Stoll, J. (2023, November). *Number and share of live sports viewers in the United States from 2021 to 2025*. Statista. https://www-statista-com.ezpv7-web-p-u01.wpi.edu/statistics/1310558/live-sports-viewers-us/

The Trustees of Princeton University. (n.d.). *Fixed effects and random effects - panel data using Stata - Research Guides at Princeton University*. Princeton University. https://libguides.princeton.edu/stata-panel-fe-re#:~:text=An%20advantage%20of%20using%20the,are%20absorbed%20by%20the%20intercept

8.0 Appendices

Appendix A: Derivation of OLS Estimator

From Equation 2, we know that $u = Y - X\beta$. Thus, we can plug this equation into $u'u$ to get $(Y - X\beta)'(Y - X\beta)$. With this equation written out, we now have to minimize β using the following minimization problem, which we can begin to simplify for solving.

$$min_{\beta} u'u = (Y - X\beta)'(Y - X\beta)$$

$$min_{\beta} u'u = (Y' - \beta'X')(Y - X\beta)$$

$$min_{\beta} u'u = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$

We can show that $\beta'X'Y = Y'X\beta$ using K as the number of independent variables and N as the number of observations. The coefficient vector β is $\mathbb{R}^{K\times1}$, so the transpose β' is $\mathbb{R}^{1\times K}$. The matrix of independent variables X is $\mathbb{R}^{N\times K}$, so the transpose X' is $\mathbb{R}^{K\times N}$. The dependent variable vector Y is $\mathbb{R}^{N\times1}$, so the transpose Y' is $\mathbb{R}^{1\times N}$. Using each matrix dimension, we can solve for the dimensions of $\beta'X'Y$ and $Y'X\beta$.

$$\beta'X'Y = (\mathbb{R}^{1\times K})(\mathbb{R}^{K\times N})(\mathbb{R}^{N\times1}) = (\mathbb{R}^{1\times N})(\mathbb{R}^{N\times1}) = (\mathbb{R}^{1\times1})$$

$$Y'X\beta = (\mathbb{R}^{1\times N})(\mathbb{R}^{N\times K})(\mathbb{R}^{K\times1}) = (\mathbb{R}^{1\times K})(\mathbb{R}^{K\times1}) = (\mathbb{R}^{1\times1})$$

As both of these terms have been proven as scalars, we know the transposition of the term is the same term. Thus, $\beta'X'Y = (\beta'X'Y)' = Y'X\beta$. Now that this has been proven, we can continue working on the minimization problem below.

$$min_{\beta} u'u = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

To solve this minimization problem, we must take the partial derivative with respect to β and set it equal to zero. We will take the partial derivative of each term for simplicity.

$$\frac{\partial Y'Y}{\partial \beta} = 0$$

$$\frac{\partial \beta'X'Y}{\partial \beta} = X'Y$$

$$\frac{\partial \beta'X'X\beta}{\partial \beta} = 2X'X\beta$$

To continue, we must prove the equality of $\frac{\partial \beta'X'X\beta}{\partial \beta} = 2X'X\beta$, which we will do below.

$X'X$ is symmetric. Let $X'X = A$

$$\beta' Ab = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_k \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{12} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

This can be simplified to:

$$\beta' A\beta = \begin{bmatrix} \sum_{i=1}^k \beta_i \cdot a_{i1} & \sum_{i=1}^k \beta_i \cdot a_{i2} & \cdots & \sum_{i=1}^k \beta_i \cdot a_{ik} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Completing this matrix multiplication results in

$$\begin{aligned} \beta' Ab = {} & \beta_1^2 \cdot a_{11} + \beta_1 \cdot \beta_2 \cdot a_{21} + \cdots + \beta_k \cdot \beta_1 \cdot a_{k1} \\ & + \beta_2 \cdot \beta_1 \cdot a_{12} + \beta_2^2 \cdot a_{22} + \cdots + \beta_k \cdot \beta_2 \cdot a_{k2} \\ & + \cdots \\ & + \beta_k \cdot \beta_1 \cdot a_{1k} + \beta_k \cdot \beta_2 \cdot a_{2k} + \cdots + \beta_k^2 \cdot a_{kk} \end{aligned}$$

With this simplified expression, we can calculate the partial derivative of $\beta'A\beta$ with respect to β.

$\frac{\partial\beta'A\beta}{\partial\beta_1} = 2\beta_1 a_{11} + \beta_2 a_{21} +...+ \beta_k a_{k1} + \beta_2 a_{12} +...+ \beta_2 a_{2k} +...+ \beta_k a_{1k} +...+ \beta_k a_{kk}$

$\frac{\partial\beta'A\beta}{\partial\beta_1} = 2(\beta_1 a_{11} + \beta_2 a_{12} +...+ \beta_k a_{1k})$

Similarly,

$\frac{\partial\beta'A\beta}{\partial\beta_2} = 2(\beta_1 a_{21} + \beta_2 a_{22} +...+ \beta_k a_{2k})$

…

$\frac{\partial\beta'A\beta}{\partial\beta_k} = 2(\beta_1 a_{k1} + \beta_2 a_{k2} +...+ \beta_k a_{kk})$

Using matrix notation to simplify all of the above equations, we get

$\frac{\partial\beta'A\beta}{\partial\beta_1} = 2A\beta$

Replacing A with $X'X$, we get

$\frac{\partial\beta'A\beta}{\partial\beta_1} = 2X'X\beta$

With the proof of $\frac{\partial\beta'X'X\beta}{\partial\beta} = 2X'X\beta$ complete, we can continue with the minimization problem $min_{\beta} u'u = Y'Y - 2\beta'X'Y + \beta'X'X\beta$ using the partial derivative with respect to β (AV, 2018).

$min_{\beta} u'u = Y'Y - 2\beta'X'Y + \beta'X'X\beta$

$\frac{\partial(u'u)}{\partial\beta} =- 2X'Y + 2X'X\beta = 0$

$X'X\beta = X'Y$

$\beta = (X'X)^{-1}X'Y$

With this result, we have solved for the least squares estimator β in the OLS regression model (AV, 2015).

Appendix B: Multiple Linear Regression Python Code

```
# Import necessary libraries
import pandas as pd
import statsmodels.api as sm

# Read the csv into a Pandas dataframe
data = pd.read_csv('data.csv', encoding='Windows-1252')

# Define X and y
X = data[['pick', 'gm_exp', 'nfl_pl_exp', 'nfl_exec_exp', 'nfl_hc_exp',
'ncaa_hc_exp’]]
y = data['pff_grade']

# Add a constant term to the predictor variables
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X).fit()

# Print the summary of the regression
print(model.summary())
```

Appendix C: Time Fixed Effects Regression Python Code

```
# Import necessary libraries
import pandas as pd
import numpy as np
import statsmodels.api as sm
from linearmodels.panel import PanelOLS

# Read the csv into a Pandas dataframe
data = pd.read_csv('data.csv', encoding='Windows-1252')

# Set the index for panel data
data.set_index(['player_code', 'year'], inplace=True)

# Fit the fixed effects model and print results
mod = PanelOLS(data['pff_grade'], data[['pick', 'gm_exp', 'nfl_exec_exp',
‘nfl_pl_exp’]], entity_effects=False, time_effects=True)
res = mod.fit(cov_type='clustered', cluster_entity=True)
print(res)
```

Appendix D: Mixed Effects Regression Python Code

```
# Import necessary libraries
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Read the csv into a Pandas dataframe
data = pd.read_csv('data.csv', encoding='Windows-1252')

# Fit the mixed effects model and print results
model = smf.mixedlm("pff_grade ~ pick + gm_exp + nfl_exec_exp + nfl_pl_exp
+ C(year)", data, groups=data["player_code"])
result = model.fit()
print(result.summary())
```