

# College Football Predictions



## Project Milestone 5

Tyon Davis, Victoria Ramirez-Garcia, Cameron Davis, Kairee Gay, Kiana Forbes

December 5, 2023

## Table of Contents

Project Milestone 5 .....	1
Writeup .....	3
[10%] Intuition - why should it be better than the state of the art?.....	3
[35%] Description of your approaches: algorithms, user interfaces, etc. ....	3
[2%] Introduction - Motivation.....	5
[3%] Problem Definition in Paragraph Form.....	5
• [5%] Description of your testbed; list of questions your experiments are designed to answer.....	6
[25%] Details of the experiments; observations .....	11
[5%] Conclusions and discussion.....	16
Survey .....	17
[5%] Survey .....	17
Distribution Of Team Member Effort.....	28
[-5% if not included] Distribution of team member effort. Can be as simple as "all team member contributes a similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who contributed more .....	28
Reference .....	29

## Writeup

### [10%] Intuition - why should it be better than the state of the art?

The football analytics program aims to surpass the current state of the art in football analytics by offering a unique combination of advanced features. The comprehensive range of defensive metrics, including Solo Tackles, Assisted Tackles, Total Tackles, Sacks, and more, provides a holistic view of team performance. The program introduces interactive visualizations, such as stacked bar charts and scatter charts with football symbols, offering a dynamic and engaging user experience for exploring defensive statistics. The inclusion of a slider for threshold-based filtering adds a tailored perspective, allowing users to focus on teams that meet or surpass specific performance thresholds in selected defensive categories. The radar chart facilitates team comparison, offering a simultaneous assessment of defensive performance for two selected teams. Additionally, hidden charts, external links, strategic insights, injury risk analysis, and predictive aspects enhance the program's depth and functionality, making it a comprehensive solution for football analytics.


### [35%] Description of your approaches: algorithms, user interfaces, etc.

- update data

The provided Python program is designed for web scraping college football player statistics from ESPN's website. It employs a modular structure, utilizing various external functions and modules such as `setDriver`, `UpdateCSV`, `scrape_header`, `week_finder`, and `get_url`. The main function, `scrape_players_and_stats`, takes advantage of the BeautifulSoup library for HTML parsing and Selenium for automated browser interactions. Notable patterns include the handling of dynamic content through a while loop that clicks the "Show More" button until it's no longer available. The code demonstrates a systematic approach to collecting player names, schools, and statistics, grouping the latter by headers and saving the data to CSV files based on the URL. Additionally, the `scrape_team_future_info` function is designed to extract future team information, including details like away team, home team, time, TV information, and ticket details. The use of conditional statements helps determine the type of statistics being scraped based on the URL, directing the data to the appropriate CSV file. Overall, the program showcases efficient web scraping practices, organized code structure, and automation for data collection from dynamic web pages.

- create charts

The football analytics dashboard sources its data from a CSV file named "defense.csv," encompassing diverse defensive metrics across various college football teams. With a visually engaging dark-themed layout, the dashboard features a navigation bar housing links to "About" and "Contact" sections. The main dashboard section, titled "2023 Football Defensive Stats," showcases a variety of charts for visualizing defensive statistics.

Among the primary charts, the stacked bar chart (id='allD-chart') presents aggregated defensive statistics for different college football teams, with interactive capabilities for potential data filtering based on the x-axis range. The scatter chart (id='football-chart') employs football symbols ('' ) to represent each team, symbol size dynamically reflecting corresponding defensive statistic values. The radar chart (id='radar-chart') facilitates the comparison of selected defensive metrics for two chosen teams through polygons, illustrating performance across various defensive categories.

The dashboard offers interactive controls, including dropdowns and sliders for users to select defensive statistic categories and compare teams. A slider enables users to adjust a threshold value, filtering teams based on the chosen defensive statistic. Team selection dropdowns empower users to choose two teams for direct defensive performance comparison.

Additionally, the application incorporates hidden charts (Passing, Running, Sack, Score, Total, Total Histogram) that become visible upon users selecting two teams, providing detailed insights into specific facets of the game.

Thematically, the application employs football-inspired colors, utilizing green for the background, yellow for text, and random colors for charts. The navigation bar includes external links to web pages offering additional information under "About" and "Contact."

- Team Strengths and Weaknesses:

Identify specific defensive areas where teams excel or struggle, offering crucial insights into their strengths and weaknesses.

- Player Performance:

Evaluate individual player contributions to leading Offensive statistics, enabling teams and coaches to recognize standout performers and areas for improvement.

- Tactical Adjustments:

Coaches can use the detailed defensive metrics to make informed tactical adjustments, optimizing strategies based on the strengths and weaknesses revealed in the data.



- Opponent Analysis:

Gain a competitive edge by studying the defensive performance of upcoming opponents. Teams can tailor their offensive strategies to exploit identified weaknesses in the opposition's defense.

- Game Predictions:

Leverage historical defensive data to make predictions about future games. Analyzing trends and patterns can contribute to more accurate forecasts of team performance.

- Fan Engagement:

Enhance fan engagement by providing visual and easily understandable defensive statistics. Fans can gain deeper insights into their favorite team's performance and engage in data-driven discussions.

- Overall Team Performance:

Evaluate the overall defensive performance of a team across multiple categories. This information is valuable for assessing a team's competitiveness and potential success in the league.

- Scouting Opponents:

Scout opposing teams to understand their defensive strategies and key players. This knowledge can be crucial in developing effective offensive game plans.

## [2%] Introduction - Motivation

The motivation behind this project stems from the need for a more reliable and data-driven approach to predicting the future performance of college football players. Current methods, relying on TV sports analysts and individual assessments from sports websites like ESPN, present limitations in terms of subjectivity, emotional biases, and a focus on popular players and teams. By leveraging advanced data analytics, we aim to develop a program that provides accurate predictions, considering various performance metrics, player-specific characteristics, and game-specific factors. This initiative seeks to overcome the drawbacks of current practices and enhance the accuracy and inclusivity of player performance predictions in college football.

## [3%] Problem Definition in Paragraph Form

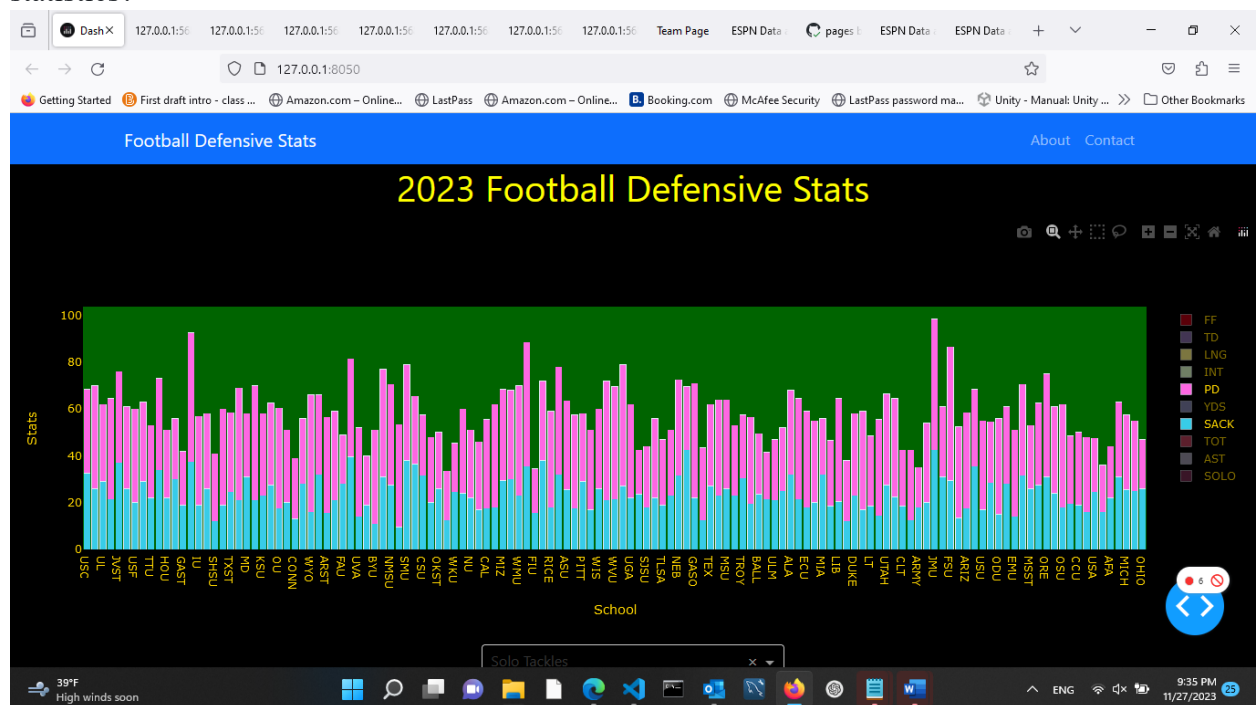
The challenge at hand is to create a predictive model for college football player performance in upcoming games. Current approaches, involving TV sports analysts and individual assessments on sports websites, are hindered by subjective biases, emotional influences, and a tendency to focus on high-profile players and teams. The goal is to move beyond these limitations by developing a program that relies on historical game data, incorporating diverse performance metrics and player-specific attributes. This

model will account for the unique characteristics of each player and the specific conditions of the upcoming game to deliver more accurate and unbiased predictions. The aim is to revolutionize the way we forecast player performances, providing a data-driven and comprehensive solution for college football enthusiasts.

- [5%] Description of your testbed; list of questions your experiments are designed to answer

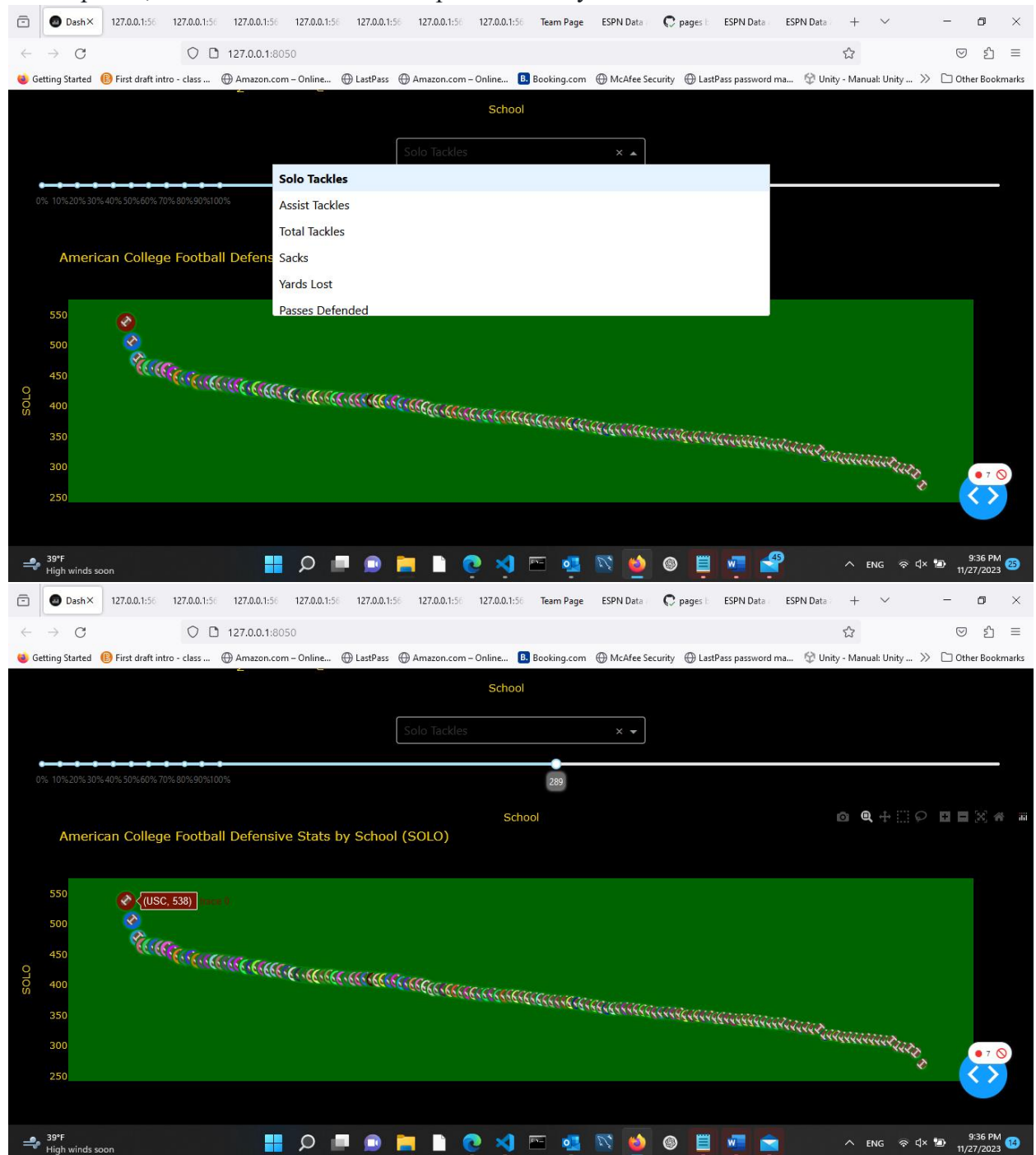
- Overall Defensive Performance:

Which teams have the strongest overall defensive performance based on aggregated statistics?

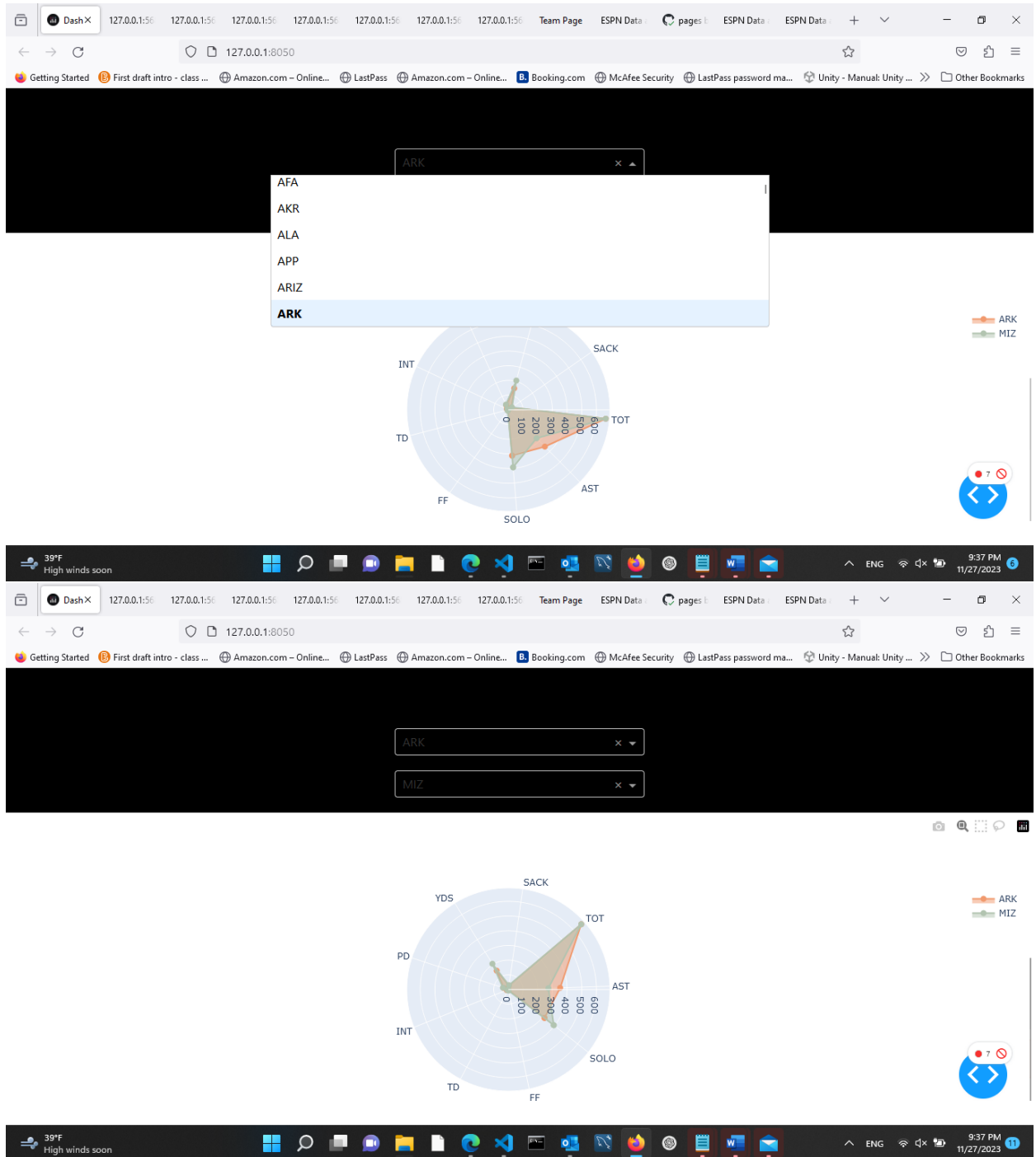


- Category-specific Insights:

What teams excel in specific defensive categories, such as Solo Tackles, Sacks, Interceptions, or Forced Fumbles? -Comparative Analysis:



How do two selected teams compare in terms of overall defensive performance and specific defensive categories?





Which team is a running team or passing team? Is the team balanced with running and passing stats?



- Opponent Vulnerabilities:  
What defensive weaknesses do upcoming opponents exhibit, and how can these be exploited strategically?



- **Threshold Analysis:**

Which teams meet or exceed a user-defined threshold for a specific defensive category, and how does this impact their overall performance?



- **Game-specific Insights:**

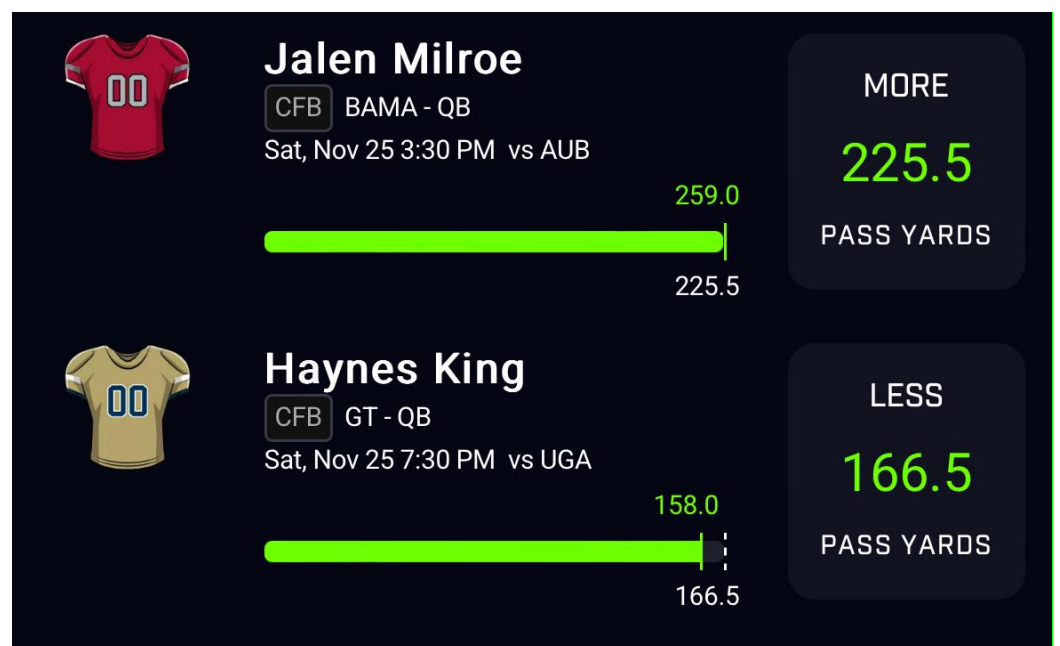
How does a team's RB & QB offensive performance vary from game to game, and are there patterns related to specific opponents?

- **Fan Engagement:**

What defensive statistics are most engaging for fans, and how can the program enhance fan understanding and interaction with the team's performance?

### [25%] Details of the experiments; observations

During our experimental sessions on PrizePicks, where users utilized our program in groups of two players specializing as quarterbacks (QB), running backs (RB), or wide receivers (WR), we implemented a betting strategy based on the numerical values provided by PrizePicks. Users were able to make predictions on whether the actual scores would be higher or lower than the given numbers. The graphical representation of our outcomes showed green lines for correct predictions and red lines for incorrect ones. Remarkably, out of the 16 players assessed, users employing our program achieved accurate predictions for 13 instances. This success rate suggests that our program demonstrated effectiveness in assisting users in making informed decisions. Notably, our observations indicated that PrizePicks often set scores within a range where outcomes could swing either way for certain games. This aligns with our approach, showcasing that users, with the assistance of our program, were able to capture the unpredictable nature of the scores.





## Jayden Daniels

CFB LSU - QB

Sat, Nov 25 12:00 PM vs TXAM

235.0



LESS

320.5

PASS YARDS



## Payton Thorne

CFB AUB - QB

Sat, Nov 25 3:30 PM vs BAMA

91.0



LESS

155.5

PASS YARDS



## J.J. McCarthy

CFB MICH - QB

Sat, Nov 25 12:14 PM vs OSU

148.0



LESS

201.5

PASS YARDS



## Kyle McCord

CFB OSU - QB

Sat, Nov 25 12:14 PM vs MICH

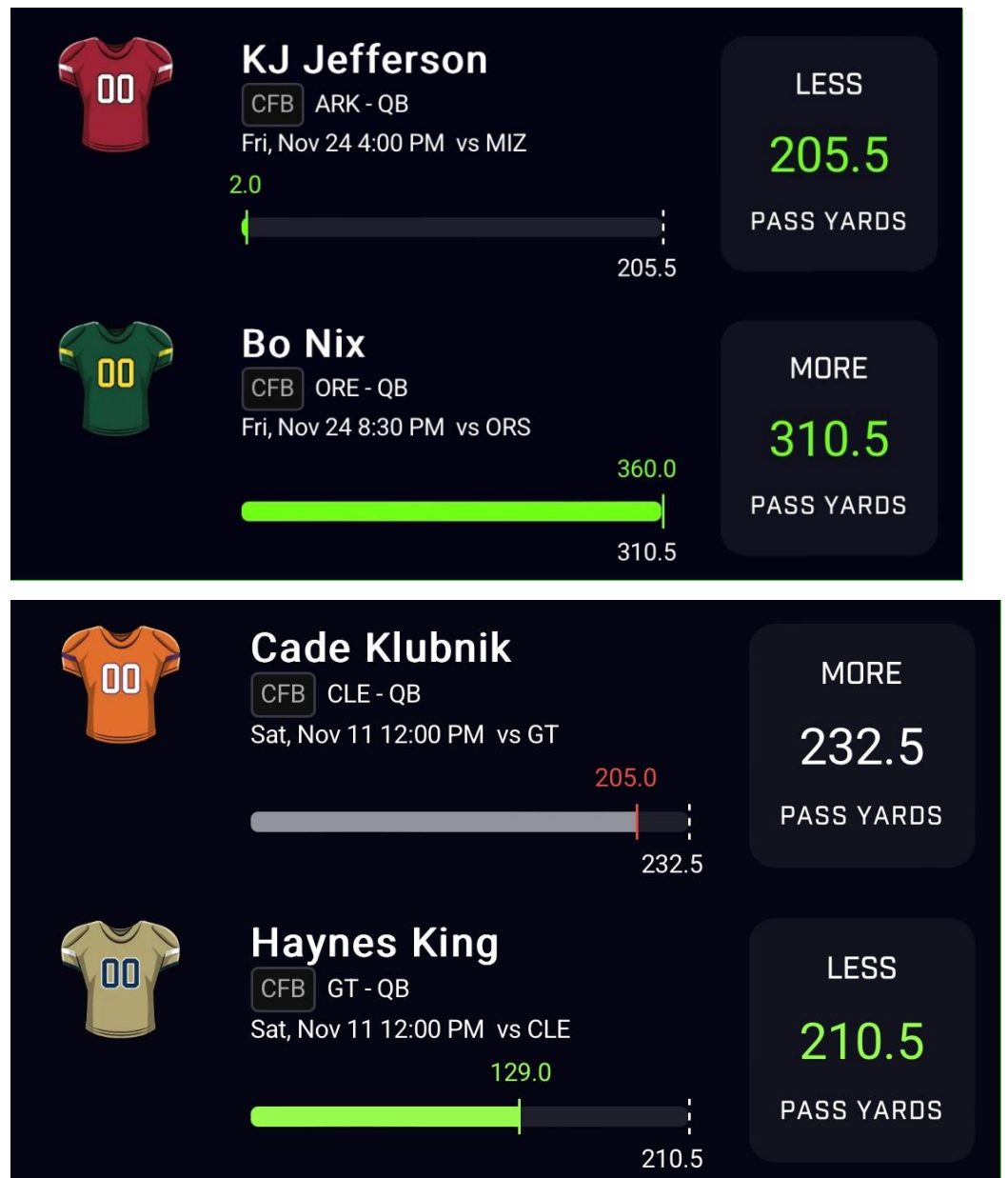
271.0



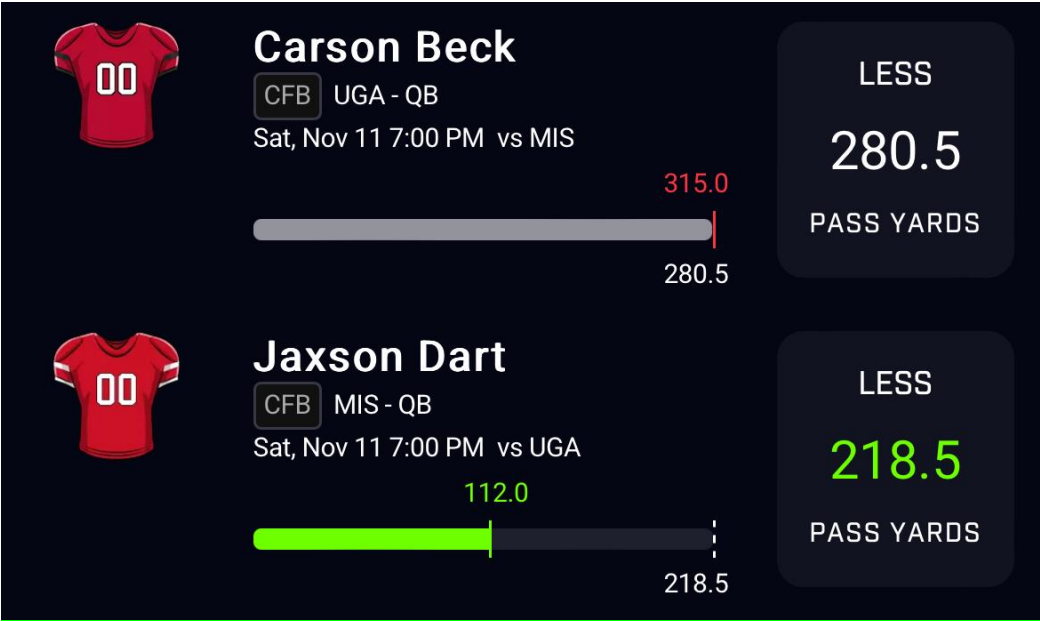
LESS

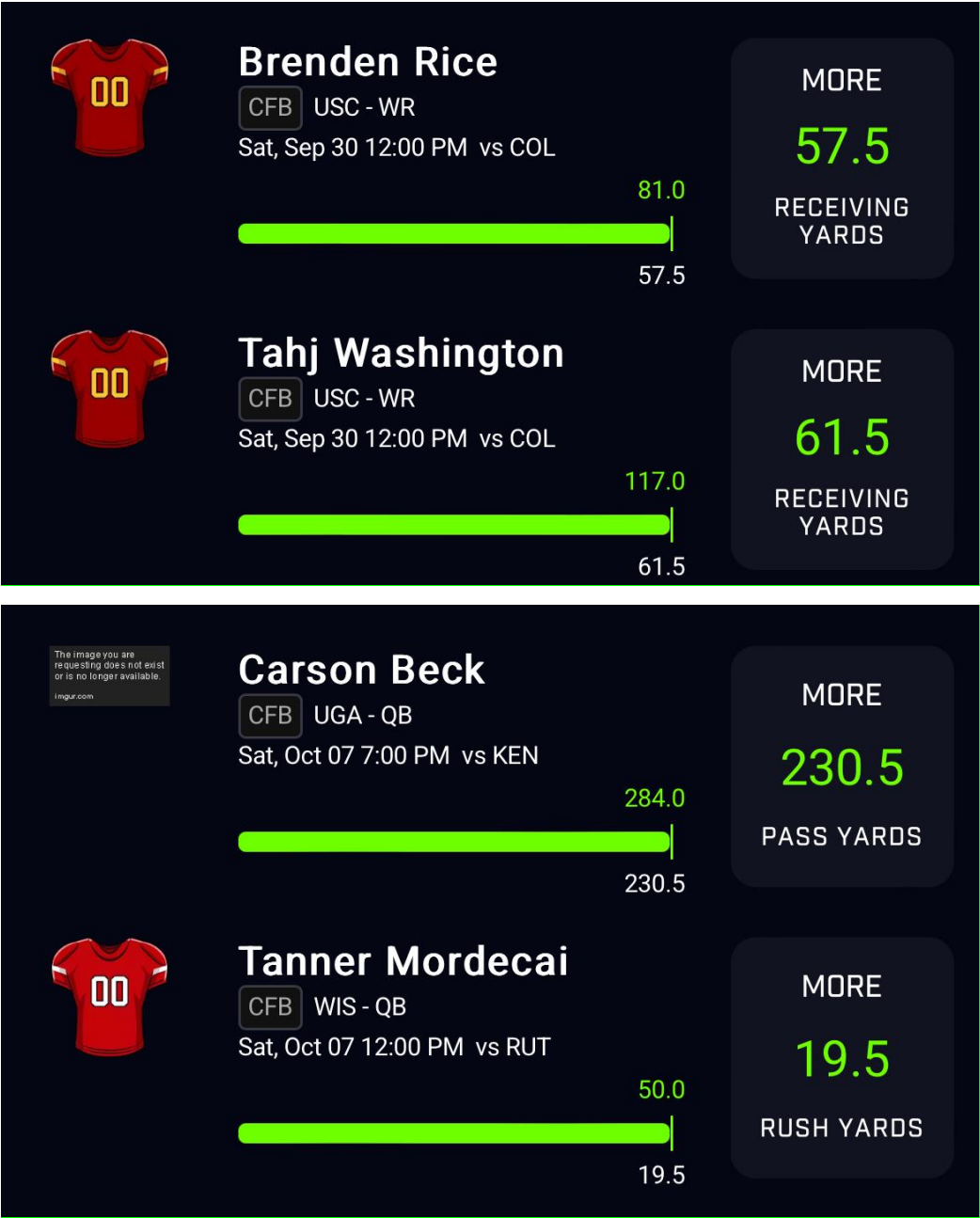
237.5

PASS YARDS









## [5%] Conclusions and discussion

In conclusion, for those interested in furthering this project, a logical next step would be to enhance the program's capabilities to autonomously generate predictions for users. Our experience highlighted the challenges associated with web scraping, with messy data retrieval being a common hurdle. Leveraging tools like Selenium proved invaluable for programmers in streamlining the process. Once the data was successfully scraped and stored in CSV files, the subsequent analysis became significantly more manageable.

From the collected data, valuable insights emerged, particularly regarding the Total (TOT) chart. It was observed that the TOT chart, resembling a bell curve, provides a useful indicator for predicting potential winners. Teams positioned to the left with lower TOT values were found to have a higher likelihood of winning. This observation implies that TOT serves as a significant predictor of game outcomes.

Overall, our findings suggest that integrating the program into the prediction process enhances accuracy for users. However, to further improve the program, programmers should consider incorporating additional variables that account for scenarios where score differentials become substantial, or key players are benched, impacting overall team performance. It's crucial to acknowledge that while predictions contribute valuable insights, the inherent unpredictability of sports, including off days for teams, implies that correctness cannot be guaranteed in every instance.

## Survey

[5%] Survey

**Link:**

<https://blog.collegefootballdata.com/predicting-spreads-gbdt/>

**(a) Main Idea:**

The main idea of the blog post titled "Predicting Spreads with Gradient Boosted Decision Trees (GBDT)" is to explore the application of machine learning techniques, specifically Gradient Boosted Decision Trees, to predict point spreads in college football games. In this post, I found that the author demonstrates how historical game data, player statistics, and other factors can be used to build a predictive model for determining point spreads in college football matchups.

**(b) Usefulness for My Project: This blog post could be useful for my project in several ways:**

**Machine Learning Techniques:** It introduces me to the use of machine learning techniques, specifically Gradient Boosted Decision Trees, which I might find applicable to my project when creating predictive models for player performance.

**Predictive Modeling:** It provides insights into how predictive modeling can be applied to college football-related data. While the blog post focuses on point spreads, I can adapt similar techniques for player performance forecasting.

**(c) Potential Shortcomings and Improvements:**

**Data Quality:** The quality and cleanliness of the data used in predictive modeling are critical. I need to ensure that the data I collect for player performance is clean and free from errors, as data quality can significantly impact model accuracy.

**Link:**

<https://thepowerrank.com/2018/08/14/how-to-make-accurate-football-predictions-with-linear-regression/>

**(a) Main Idea:**

The main idea of the blog post titled "How to Make Accurate Football Predictions with Linear Regression" is to explain the application of linear regression, a fundamental statistical technique, in making accurate football predictions. In this post, I found that the author demonstrates how linear regression can be used to analyze football-related data, derive meaningful insights, and make predictions about football outcomes, such as point spreads.

**(b) Usefulness for My Project:**

This blog post is highly relevant and useful for my project for several reasons:

**Statistical Technique:** It provides a detailed explanation of using linear regression, a statistical technique, for football predictions. I can consider incorporating linear regression into my project's modeling approaches for forecasting player performance or game outcomes.

**Data Analysis:** The blog post emphasizes the importance of data analysis and the use of relevant features in predictive modeling. This aligns with my project's goal of collecting and analyzing data on college football players' performances.

**(c) Potential Shortcomings and Improvements:**

While the blog post offers valuable insights, I should be aware of potential shortcomings and consider areas for improvement in my project:

**Model Evaluation:** I should explore various model evaluation techniques to assess the performance of my predictive models effectively.



**Link:**

<https://www.samford.edu/sports-analytics/fans/2022/The-Slow-Growth-of-Data-Analytics-in-the-Football>

**(a) Main Idea:**

The main idea of the article titled "The Slow Growth of Data Analytics in Football" is to discuss the relatively slow adoption and integration of data analytics within the football industry, particularly at the college and NFL levels. The article highlights various challenges and barriers that have hindered the widespread use of data analytics in football, such as resistance to change, lack of data infrastructure, and cultural factors. It also emphasizes the potential benefits that data analytics can bring to sport.

**(b) Usefulness for My Project:**

It offers an understanding of the challenges and hurdles faced by the football industry when it comes to adopting data analytics. This awareness can help me anticipate potential challenges in my own project.

Identifying Opportunities: By highlighting the slow growth of data analytics, the article indirectly points to opportunities for improvement and innovation in this field. It encourages me to think about how my project can contribute to the advancement of analytics in football.

Cultural Considerations: The article touches on cultural factors that influence the acceptance of analytics in football. This can be relevant when considering how stakeholders in the sport may perceive or resist the use of data in player performance forecasting.

**(c) Potential Shortcomings and Improvements:**

Depth of Analysis: The article provides a broad overview of the challenges and benefits of data analytics in football. To apply these insights effectively to my project, I may need to delve deeper into specific aspects, such as data sources and modeling techniques.

Timeliness: The article's information may be based on the state of the industry at the time of publication. I should ensure that I consider the most recent developments and trends in data analytics in football.

**Link:**

<https://mcillecesports.com/>

**(a) Main Idea:**

The main idea of this page is to introduce and promote the "2023 College Football Atlas." The atlas is described as highly detailed and comprehensive analytics preview of the college football season, covering various aspects of the game, including predictions, statistics, coaching, and more. It emphasizes the importance of adapting prediction methods to account for the changing landscape of college football, particularly the impact of the Transfer Portal.

**(b) Usefulness for My Project:**

It highlights the importance of adapting analytics and prediction methods to changes in the sport, which is a valuable reminder for anyone involved in sports analytics projects. It shows me that not all players would have had stats with the team from previous year.

**(c) Potential Shortcomings and Improvements:**

This site would not be one where I can gather potential shortcomings and improvements from. Well, no improvements that I feel should be enforced.

**Link:**

<https://www.sciencedirect.com/science/article/pii/S1877050914011181>

**a) Main idea:**

The main idea of this paper named “Sports data mining: predicting results for the college football games,” is that there are different approaches that everyone takes when it comes to predicting these games with simple methods.

Here it is represented a new method of predicting by using a data mining approach to get more accurate results. Data mining has methods that are more quantitative than subjective.

**b) Why it will be useful for your project:**

This paper could be useful because it collects different data showing statistics in a different view than the normal human predictions. Rather than viewing the predictions of a professional predictor, we can evaluate the method that the authors of the paper used, which was sports data mining. We can also look at the different statistics being solved.

**c) its potential shortcomings, that you will try to improve upon:**

Although the information of this paper includes new methods that could potentially give more accurate predictions, the data mining approach could be overcomplicated, seeing as to there must be many different equations.

Depending on the kind of statistics one is looking for, it can get lost in translation. This also does include older data, not the most recent years.

**Link:**

[https://journals.lww.com/nsca-jscr/Fulltext/2019/01000/National\\_Scouting\\_Combine\\_Scores\\_as\\_Performance.12.aspx](https://journals.lww.com/nsca-jscr/Fulltext/2019/01000/National_Scouting_Combine_Scores_as_Performance.12.aspx)

**a) Main idea:**

The main idea of this paper is to show how the statistics of college players' records compares to how well their performance was when they were invited to the National Scouting Combine, that led them to be selected in a draft. Every player, depending on their playing position, performed a variety of tests that were then compared to their on-field performance.

**b) Why it will be useful for your project:**

We want to predict the performance of college football players and this paper gives results of players who play different positions. It takes the linear regressions and combines the results to view how the predictive results were.

It also shows the results of the different positions and the outcomes of the players that were selected in the end. This is an older paper; therefore, it could be used to compare new statistics with the formulas.

**c) Its potential shortcomings, that you will try to improve upon:**

Potential shortcomings: The results were taken from subjects of 2005-2019 that could potentially mean that the difference of how players were in those times, have changed in training methods amongst other things.

**Link:**

<https://www.amazon.com/Graphs-Networks-Sandra-Kingan/dp/111893718X>

**a) Main idea**

SEP Graphs and Networks: Multilevel Modeling" provides insights and techniques for analyzing complex networks and graphs with a focus on understanding their hierarchy. Using the techniques discussed in the book, we can build predictive models for player performance in college football. This may involve analyzing historical player data, such as statistics, injury history, and playing positions, to predict future performance.

**b) why (or why not) it will be useful for your project.**

By leveraging network analysis, one can identify influential players and teams within the football ecosystem, while multilevel modeling helps account for hierarchical data structures, improving the accuracy of predictions. These methods can be applied to player performance prediction, game outcome forecasting, ultimately leading to more informed decisions in college football predictions and analytics.

**c) its potential shortcomings, that you will try to improve upon.**

This is a high-level mathematics book that does not have a clear path of implementation of graphs and networks of statistical data. We will have to clearly understand the rigor of these formulas before we can properly use them.



**Link:**

<https://www.liveabout.com/football-101-the-basics-of-football-1333784>

**(a) Main Idea:**

The main idea of the article titled "Football 101: The Basics of Football" is to provide an introductory overview of the fundamental aspects of American football. The article aims to educate newcomers to the sport by explaining key concepts such as the field, positions, basic rules, scoring, and terminology. It serves as a beginner's guide to understanding the basics of football.

**(b) Usefulness for My Project:**

While the article focuses on explaining the fundamentals of football, it may not directly relate to my project of gathering data on college football players' performances for future forecasting. However, it could be useful for individuals who are new to football and need to understand the sport's basics before diving into more advanced aspects, such as player performance analysis.

**(c) Potential Shortcomings and Improvements:**

The article is designed as an introductory guide, so it does not delve into advanced or analytical aspects of football. Therefore, it may not offer in-depth insights for sports analytics projects like mine. However, it serves its purpose well as an educational resource for those new to the sport.

**Link:**

<https://www.sportsinfosolutions.com/2022/04/27/study-college-football-injury-overview/#:~:text=Below%20are%20the%2025%20most-injured%20teams%20in%20the,all%20in-game%20injuries%20tagged%20and%20off-field%20injuries%20reported.>

**(a) Main Idea Recap:**

The main idea of this article is to analyze injuries in college football, especially with an eye on their impact on NFL Draft prospects. The article introduces a color-coded system with green, yellow, and red flags to evaluate the injury risk associated with players. This system is used as a reference point during NFL Draft considerations.

**(b) Usefulness for Our Project Reminder:**

While we are primarily focused on gathering data for college football player performance forecasting, understanding injuries is essential. Injuries can significantly influence a player's performance and, therefore, our forecasting models. By studying injury trends, we gain insights that can indirectly enhance the accuracy of our predictions.

**Key Takeaways :**

**Injury Risk Flags:** We've learned about the color-coded flags used to categorize injury risk for NFL Draft prospects. These flags help teams assess the potential impact of a player's injury history on their professional career.

**Positional Trends:** The article highlights how different positions in football have varying levels of injury risk. For instance, defensive linemen face a higher risk of injury, especially concerning shoulder injuries affecting the rotator cuff.

**Body Parts and Injuries:** Knee and ankle injuries were the most common during the 2021 season. We also discussed how ACL tears can be a pivotal factor in a player's draft position.

**Injury Designations:** The article provides a list of players with red and yellow flags, as well as those who are currently injured. These designations carry significant weight in NFL Draft discussions.

**(c) Potential Shortcomings and Improvements:**

It's essential to consider that injury data, while informative, may be just one aspect of player performance forecasting. Combining injury data with performance metrics and other factors can lead to more accurate predictions.

**Link:**

<https://www.actionnetwork.com/ncaaf/how-to-bet-on-college-football-using-advanced-stats-to-handicap-games>

**Main Idea:**

The primary idea of this article is to provide insights into how advanced statistics and data analysis can be used to make informed decisions when betting on college football games. It emphasizes the importance of understanding and utilizing data-driven strategies to increase the chances of success in sports betting.

**Usefulness for Our Project:**

While our main project centers around gathering data on college football player performances for forecasting, this article introduces a different perspective. It highlights the role of data and statistics in the broader realm of college football, particularly in sports betting. Understanding how data is used to analyze and predict game outcomes can offer valuable insights into the sport's intricacies.

**Potential Shortcomings:**

**Overemphasis on Data:** One potential shortcoming is an overemphasis on data-driven strategies. While data is a powerful tool, it may not account for unexpected events, like player injuries during a game or emotional factors that can influence team performance.

**Complexity for Beginners:** The article assumes a certain level of familiarity with sports betting and advanced statistics. It might not be beginner-friendly, potentially excluding those new to sports betting who would benefit from a more basic introduction.

**Limited Scope:** The article mainly focuses on advanced statistics and data analysis but doesn't delve deeply into other important aspects of sports betting, such as understanding betting odds, strategies for different bet types (e.g., spread betting, moneyline), and bankroll management.

## Distribution Of Team Member Effort

[-5% if not included] Distribution of team member effort. Can be as simple as "all team member contributes a similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who contributed more

Kiana Forbes, Cameron Davis, Victoria Rmirez-Garcia	Designer
Tyon Davis, Kairee Gay	Coders
Tyon Davis, Victoria Rmirez-Garcia, Kairee Gay, Kiana Forbes, Cameron Davis	Researcher
Tyon Davis, Kiana Forbes, Kairee Gay	Organizer
Tyon Davis, Cameron Davis, Kiana Forbes, Kairee Gay, Victoria Rmirez-Garcia	Research survey
Tyon Davis, Cameron Davis, Kiana Forbes, Kairee Gay, Victoria Rmirez-Garcia	Time Tracker
Victoria Rmirez-Garcia, Cameron Davis, Kiana Forbes	QA tester
Tyon Davis	Pacesetter
Kairee Gay	Reviewer
Tyon Davis, Cameron Davis, Kiana Forbes, Kairee Gay Victoria Rmirez-Garcia	Heilmeier Questions
Tyon Davis, Kairee Gay	Required Components



## Reference

College Football Data. "Predicting Spreads: GB/DT." College Football Data Blog, 2023,

<https://blog.collegefootballdata.com/predicting-spreads-gbdt/>

Massey, Ed. "How to Make Accurate Football Predictions with Linear Regression." The Power Rank, 14 Aug. 2018, <https://thepowerrank.com/2018/08/14/how-to-make-accurate-football-predictions-with-linear-regression/>

Samford University Sports Analytics. "The Slow Growth of Data Analytics in Football." Samford University, 2022, <https://www.samford.edu/sports-analytics/fans/2022/The-Slow-Growth-of-Data-Analytics-in-the-Football>

McIllece Sports. <https://mcillecesports.com/>

ESPN. "Neyland Stadium - Largest College Football Stadiums." ESPN,

[https://www.espn.com/espn/thelife/photos/gallery/\\_/id/5793795/image/3/neyland-stadium-largest-college-football-stadiums](https://www.espn.com/espn/thelife/photos/gallery/_/id/5793795/image/3/neyland-stadium-largest-college-football-stadiums)

Visual Studio Code. <https://code.visualstudio.com/>

Python Software Foundation. "Python Downloads." Python.org,

<https://www.python.org/downloads/>

GitHub. <https://github.com/>

Leung, Carson K., and Kyle W. Joseph. "Sports Data Mining: Predicting Results for College Football Games." Procedia Computer Science, Elsevier, 2014,

<https://doi.org/10.1016/j.procs.2014.08.153>.

"National Scouting Combine Scores as Performance Predictors in the National Football League Draft." Journal of Strength and Conditioning Research, LWW (Wolters Kluwer), 2019,

[https://journals.lww.com/nsca-jscr/Fulltext/2019/01000/National\\_Scouting\\_Combine\\_Scores\\_as\\_Performance.12.aspx](https://journals.lww.com/nsca-jscr/Fulltext/2019/01000/National_Scouting_Combine_Scores_as_Performance.12.aspx).