Levels of Football

1. Introduction

Football is a very popular sport throughout Iowa and the United States. There are many levels of football including high school, college, and the NFL. Many fans and teams want more insight into what makes teams successful. What are the differences between the three levels? Are there any stats or metrics that can describe one level of football over the other. In this project, I will look at each level of football to see if a correlation exists throughout the game and what potentially separates them. I will use two Kaggle data sets describing college and NFL statistics, as well as Bound for statistics about Iowa High School football.

2. Data

To conduct this analysis, I will utilize multiple data sources:

- **High School Football Data:**
  I will scrape statistics from Bound[1], which is a website similar to MaxPreps. There I will find team statistics for all of the high school teams in the state of Iowa. Primarily, I want to focus on teams that are in the higher classes, because they will provide better metrics for comparison.

- **College Football Data:**
  For college football, I will use a dataset on Kaggle[2] from the 2023 season. This was the most recent season with complete statistics. I chose not to use the 2024 data due to the season still being in progress.

- **NFL Data:**
  For the NFL, I will scrape from Pro Football Reference[3], which is a website that contains NFL data from past seasons. 2023 is the season that I have chosen due to its recency and comparability across levels of football.

To ensure the highest quality of analysis, I will clean and merge the datasets vertically, focusing on common metrics such as points scored, yards per game, and winning percentage to connect the data sources.

I plan on using multiple major datasets using the three data sources above.

- **High School and College Data:**
  I will combine the high school and college data sources. This will be used to demonstrate the similarities and differences between the two levels. My hope is to provide insight into what Iowa high school teams are similar to college teams and in what way.

- **College and NFL Data:**
  I will combine the college and NFL data sources. This will be used to demonstrate the similarities and differences between the two levels. A minor goal is to provide what college

---

[1] https://www.gobound.com/
[2] https://www.kaggle.com/datasets/jeffgallini/college-football-team-stats-2019
[3] https://www.pro-football-reference.com/years/2023/index.htm

programs may be best for NFL player development outside of draft pick metrics that are often used.

- **Geographic Data:**
  Finally, I will include geographic data (state, conference) to group teams. This will consist of all the data sources. The goal is to aggregate by region and find any correlations by region, level, or both.

Table 1: Data Dictionary

Field	Type	Source	Description
team_name	Text	All	Name of team
team_level	Text	All	Level of play (HS, College, NFL)
conference	Text	College, NFL	The conference the team is in (ex. AFC)
state	Text	HS, NFL	What state the team is located in (ex. IA)
ypg	Numeric	All	Offensive yards per game
ppg	Numeric	All	Points per game
pgg_a	Numeric	All	Points per game allowed
ppg_diff	Numeric, derived	All	ppg – ppg_a
off_tds	Numeric	All	Touchdowns scored on offense
W	Numeric	All	Wins
L	Numeric	All	Losses
rush_ypg	Numeric	All	Rushing yards per game
pass_ypg	Numeric	All	Passing yards per game
rush_tds	Numeric	All	Rushing Touchdowns
pass_tds	Numeric	All	Passing touchdowns
win_pct	Numeric, derived	All	Wins/Games Played

3. Analysis

3.1 Points Scored, Allowed, and Yards per Game

The first metric that I wanted to analyze was points scored (ppg) and points allowed (ppg_a). This analysis aimed to uncover which level of football sees the most scoring and where blowouts might be more frequent. To start, I performed a Pearson correlation test between these two variables. The resulting correlation coefficient was approximately 0.84, indicating a strong positive relationship. This finding aligns with expectations—teams that score more points are likely in games where more points are allowed overall, either due to mismatches or the style of play.

To explore this further, I created a bar plot that splits the data by level of football: high school, college, and NFL (shown in Figure 1). This visualization highlights clear trends. High school and college games tend to be higher-scoring on average compared to NFL games. These levels of play are also generally more prone to significant score differentials, reflecting more frequent blowouts. On the other hand, NFL games are typically lower-scoring and more competitive, suggesting tighter defenses and a higher level of parity between teams. This aligns with the perception of the NFL as a league where games are more evenly matched.
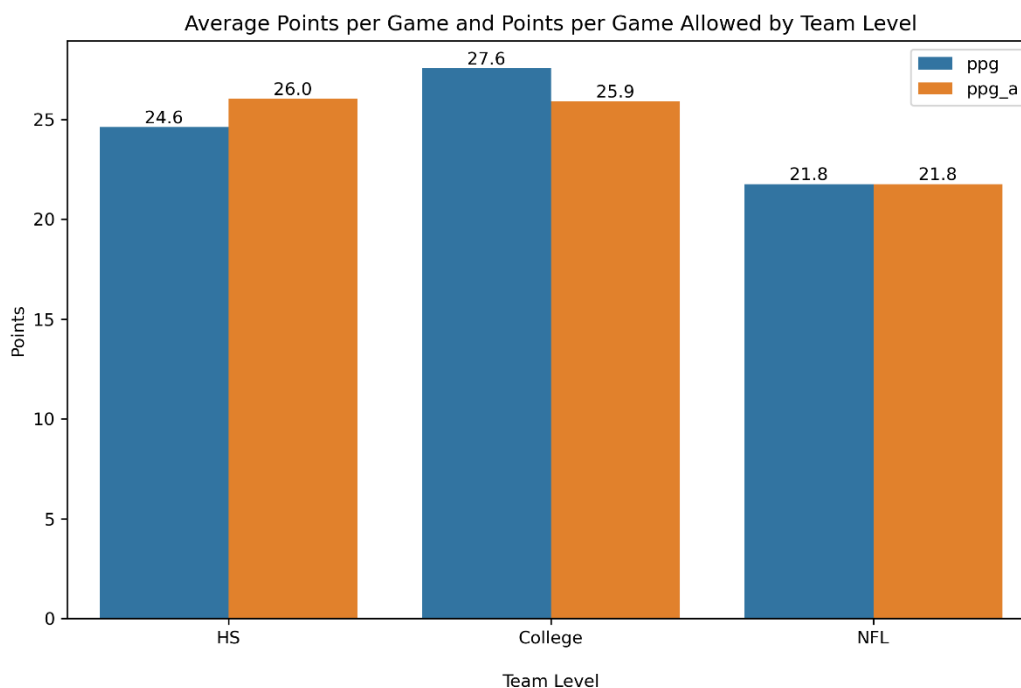

Figure 1

The next metric I analyzed was yards per game (ypg), which encompasses both passing and rushing yards to provide a comprehensive view of offensive production. Breaking this data down by level revealed notable differences between high school, college, and the NFL. High school teams average 306.9 ypg, college teams significantly outpace both other levels with an impressive 384.2 ypg, while NFL teams average 331.6 ypg.

These findings, visualized in Figure 2, underscore the fast-paced, high-octane style of play often seen in college football, which likely contributes to the higher scoring averages observed at that level. In contrast, the NFL's slightly lower ypg reflects a more balanced, strategic approach to offense, where defenses are tighter and execution is more precise. High school averages, on the other hand, may reflect the developmental nature of the game at that level, where there are wider disparities in skill and strategy.
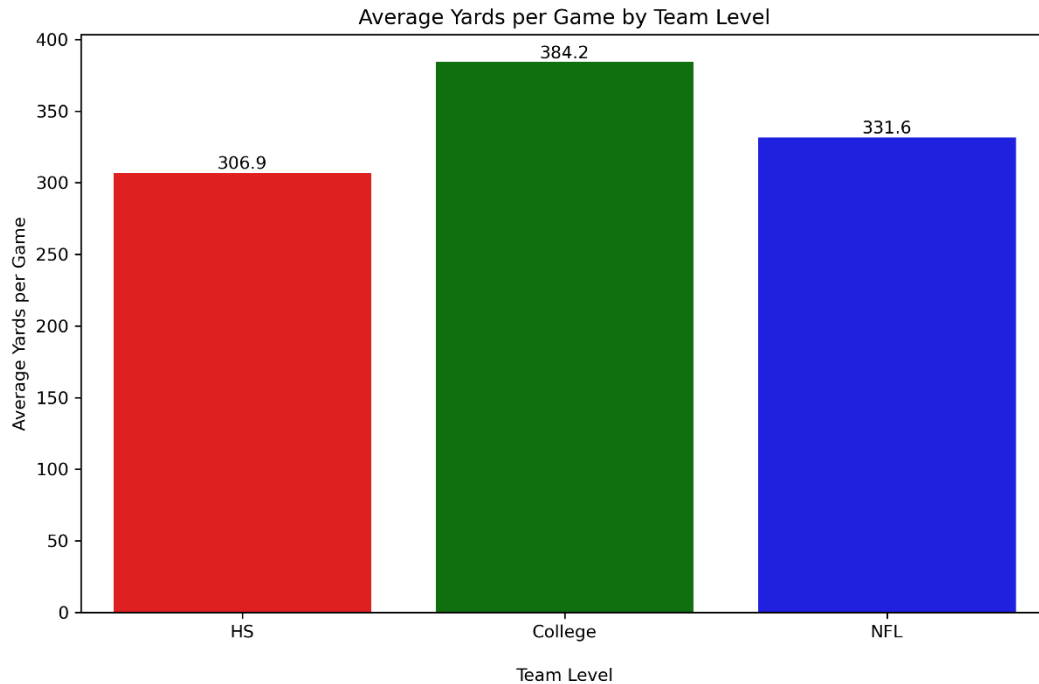
Figure 2

3.2 Win Percentage Splits

To better understand how win percentages vary across different levels and regions, I analyzed the data through a series of boxplots and bar charts.

Figure 3 displays a boxplot of win percentages by conference, showing that most conferences hover within a win percentage range of 0.4 to 0.6. Interestingly, the PAC-12 stands out with the highest median win percentage among all conferences. However, the results of an ANOVA test showed no statistically significant differences in win percentage across conferences, suggesting that these variations are not substantial enough to draw meaningful conclusions.
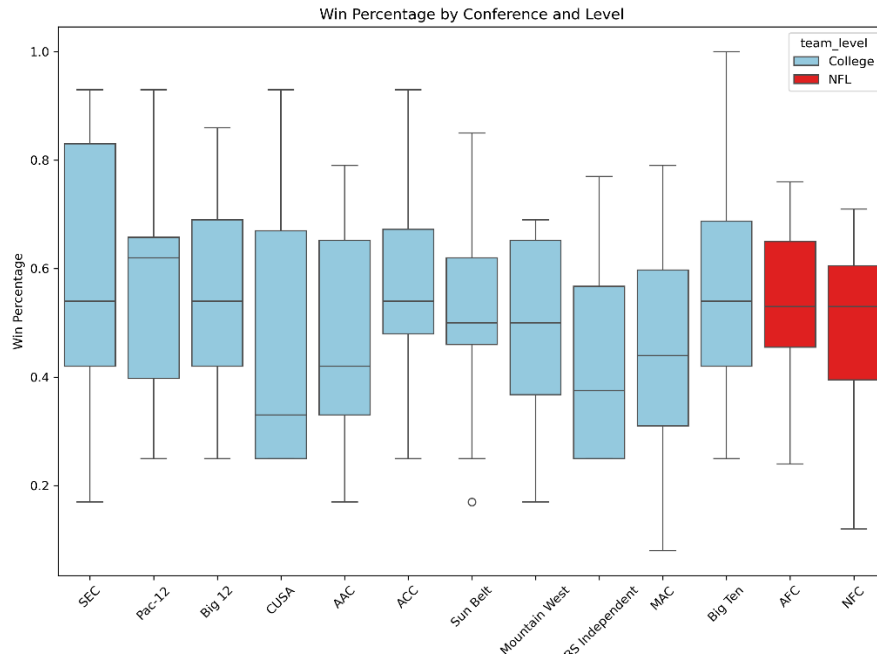
Figure 3

In Figure 4, a bar chart illustrates win percentages by region. Teams in the Southwest and West regions have the highest average win percentages at 0.51, while the Southeast lags behind with an average of 0.45. Despite these apparent differences, an ANOVA test revealed no significant variation in win percentage by region, further supporting the idea that performance is fairly consistent across regions.
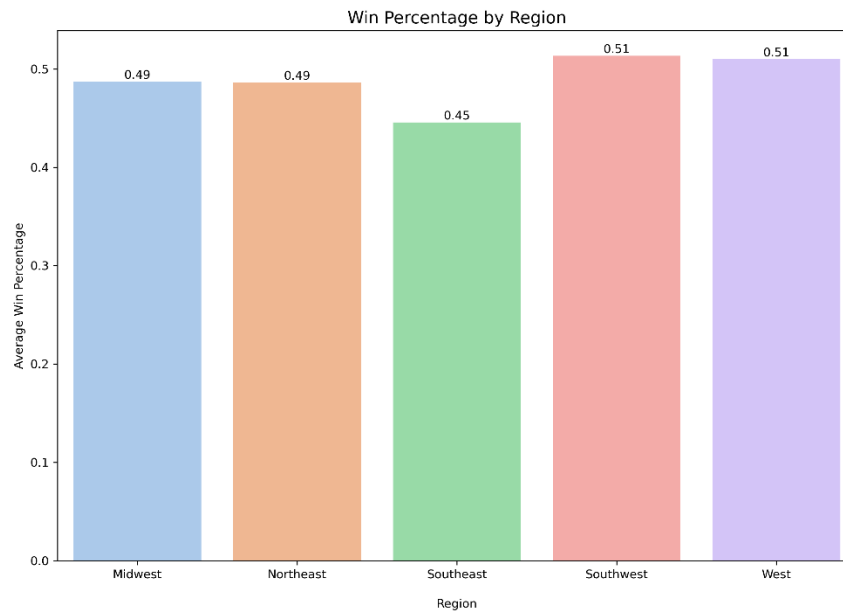

Figure 4

Finally, Figure 5 explores win percentages by state and level, focusing on high school and NFL teams. While Iowa high school teams fall in the middle of the pack—aligning with expectations—states like Michigan, New York, and Texas lead the overall win percentage rankings, with averages approaching 70%. Despite these notable trends, the ANOVA test found no significant differences in win percentage by state and level, indicating that these disparities are not statistically meaningful.
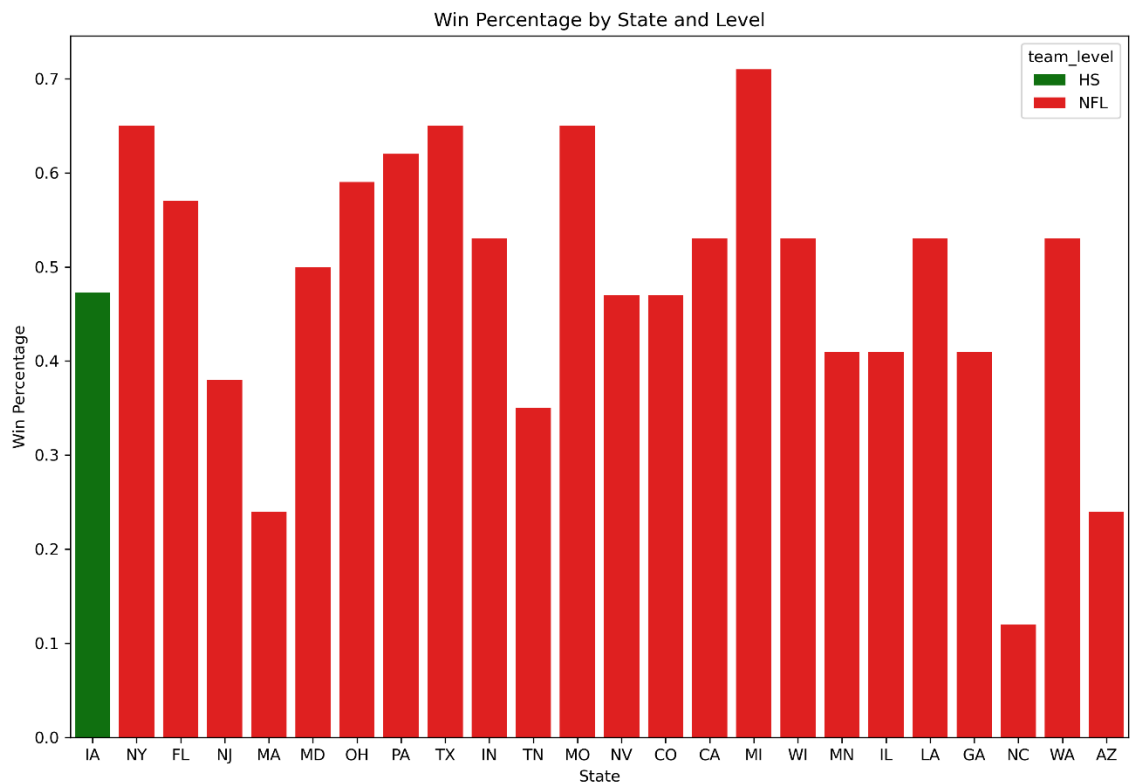

Figure 5

These findings highlight the competitive balance within football at various levels, suggesting that differences in win percentages are more influenced by specific circumstances than by overarching regional or state-level trends.

3.3 Rushing vs Passing Yards

Rushing and passing yards are fundamental metrics in football, providing insight into the offensive strategies employed by teams. To analyze this, I examined how these metrics vary by level, region, and conference.

Figure 6 compares rushing and passing yards across high school, college, and NFL teams. At the high school level, rushing yards dominate, indicating a run-heavy style of play. As teams progress to college, passing yards increase significantly, representing a major shift in offensive dynamics. From college to the NFL, the proportion of rushing to passing yards remains relatively stable, but the total yardage decreases, likely reflecting the higher level of competition and defensive efficiency. ANOVA

tests for both rushing and passing yards by level revealed statistically significant differences, confirming that the observed patterns are meaningful.
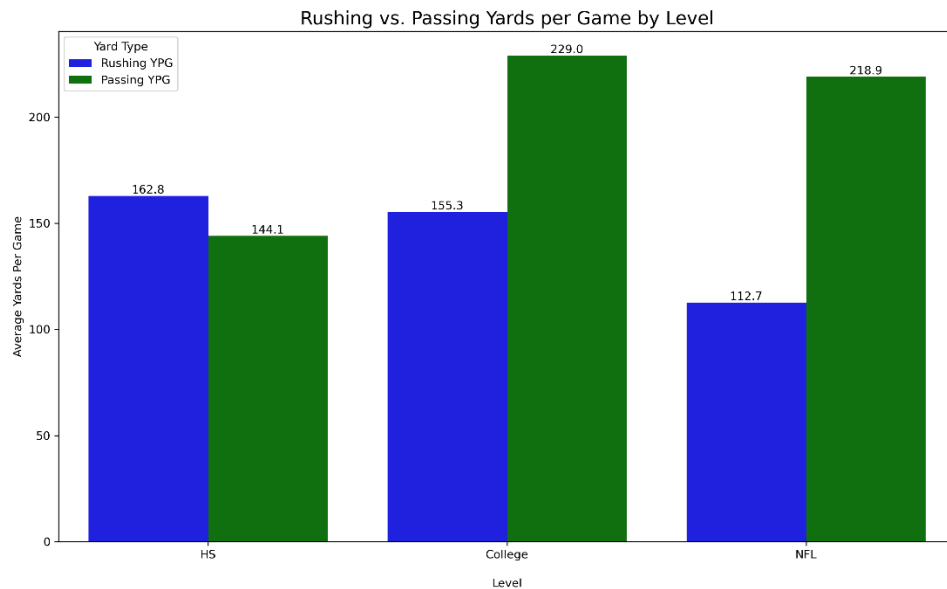

Figure 6

Figure 7 shifts the focus to regions, showing the Midwest as having the lowest average passing yards but the highest rushing yards. In contrast, the Southwest and West regions exhibit higher passing yard averages, which could be attributed to offensive schemes like Oregon's and USC's air raid systems. The ANOVA test for both rushing and passing was statistically significant, providing a useful insight into regional discrepancies.
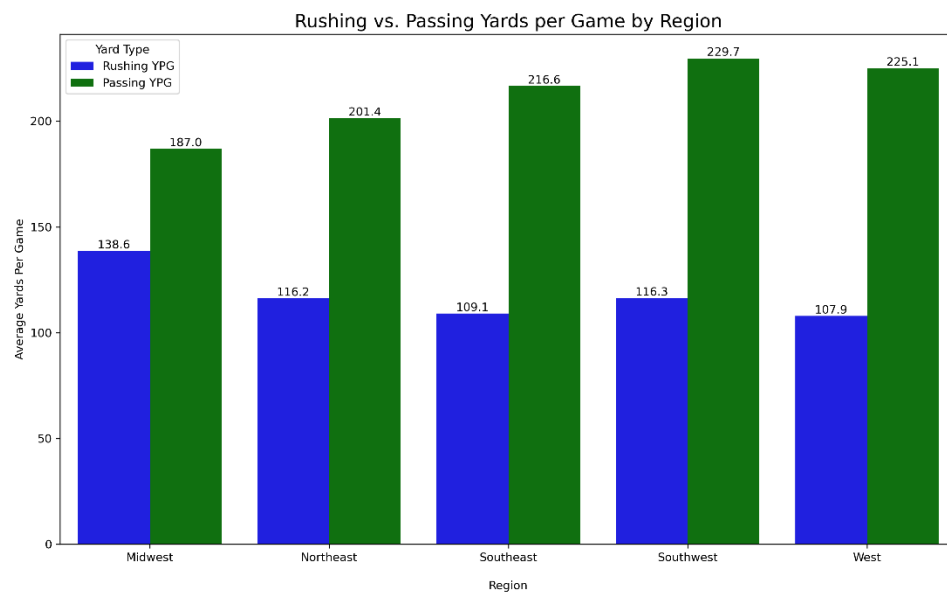

Figure 7

Figure 8 narrows the analysis further to conferences. The PAC-12 stands out with exceptionally high passing yard averages, aligning with its reputation for pass-heavy offenses. This finding is essentially a more detailed extension of the regional analysis, as conferences often align with specific geographic areas. While the PAC-12's passing emphasis is notable, the results reinforce that rushing and passing trends are highly contextual, reflecting not just the level of play but also the specific offensive philosophies of teams and regions.
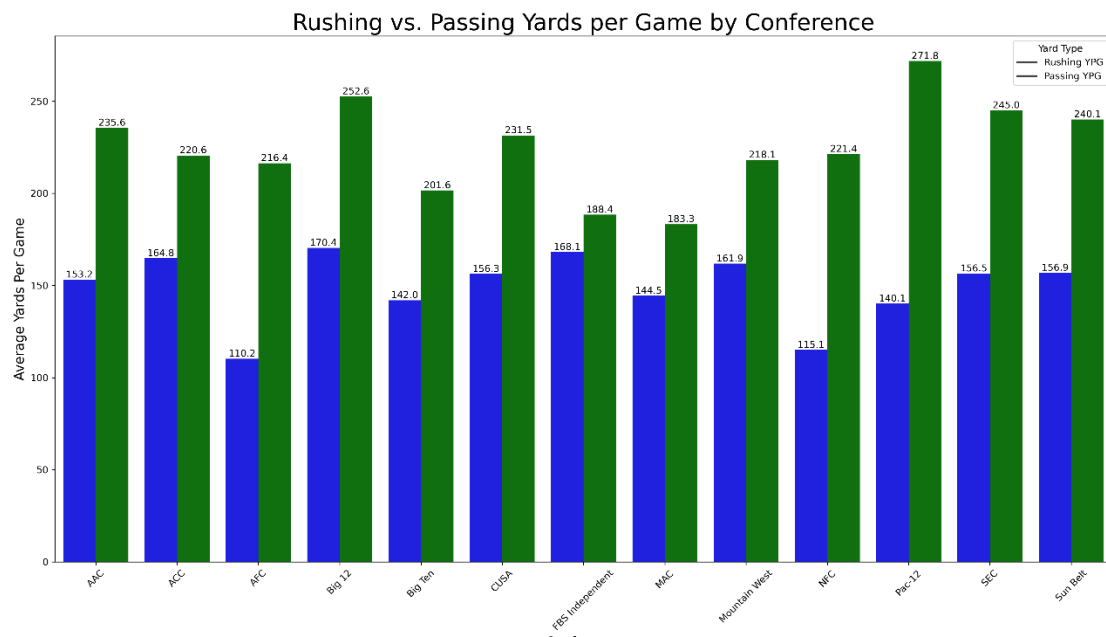

Figure 8

These observations underscore the dynamic nature of offensive strategies in football, with significant shifts in emphasis as teams move from high school to professional levels, and regional variations driven by stylistic preferences and conference identities.

3.4 Other Distinct Factors

The factors discussed previously were the major distinctions between the levels of football.

4. Conclusion

In this project, I analyzed the three levels of football: high school, college, and the NFL. The goal was to provide insights into team success and find common threads between the levels. I had 4 main research questions and one overarching question answered below:

1. *How do average scoring margins differ between levels of football?*

   The analysis revealed that scoring margins vary significantly across levels. High school and college games tend to have higher scoring margins, often reflecting a disparity in team skill levels. In contrast, NFL games are generally closer in score, as evidenced by smaller

average scoring margins. This reflects the parity and competitive balance that professional leagues strive to maintain.

2. *Is there a geographic or conference correlation (e.g. Midwest, South, etc.) that can give insight into team success?*

    Geographic and conference-level data offered mixed insights. While the PAC-12 and West/Southwest regions showed higher levels of offensive production, particularly in passing yards, the win percentage analysis revealed no significant geographic or conference correlation. Conferences generally hovered around a 40-60% win percentage, suggesting no single power 5 conference dominates in terms of overall success across the board.

3. *Do teams run or pass the ball for more yards in certain geographic regions?*

    Rushing and passing tendencies varied notably by region. The Midwest favored rushing, consistent with its reputation for traditional, run-heavy offenses. Conversely, the Southwest and West had significantly higher passing averages, likely driven by teams known for adopting innovative air-raid systems. The ANOVA test for regional differences in rushing and passing yards was statistically significant, suggesting that these trends are meaningful and reflect real regional variations in offensive strategies.

4. *In general, are there any distinct factors that can separate each level of football?*

    Yes, several factors distinguish the levels of football. College football stands out with its high-paced games, resulting in the highest average yards per game (384.2). High school games are characterized by a reliance on rushing plays and higher scoring margins, while the NFL demonstrates a balance between offensive strategies, closer game margins, and overall lower yardage totals compared to college football. These trends highlight the increased emphasis on strategy, specialization, and competitive balance as the level of play advances.

Overarching Question: What schools should a high school football player in Iowa look at?

For an Iowa high school football player, the Big Ten conference is a natural fit. The Midwest is known for producing run-heavy offenses and strong defensive teams, which align well with the skill sets often developed in Iowa high schools. Additionally, Big Ten programs like Iowa, Michigan, and Ohio St. offer competitive opportunities and pathways to professional football while maintaining football relevance.

This project has several limitations, primarily related to data availability. The analysis only includes a single year for each level of football, restricting the ability to identify trends over time. Incorporating data from multiple seasons would yield more meaningful insights. External factors such as coaching strategies, player recruitment, and budget disparities, which significantly influence team performance, were not accounted for in this study.