BAIS:3250 Final Project Report

NFL Penalties and Performance - Mady McKee

https://github.com/madylmckee/3250-final-project

1. Introduction:

In recent years, there has been growing concern among fans and analysts that the National Football League (NFL) may be "rigged" or influenced in ways that affect the fairness of games. These concerns often stem from controversial calls and penalties that appear to disproportionately benefit certain teams. Penalties can significantly affect the momentum of a game, such as halting offensive drives, granting opponents strategic advantages, or changing the overall outcome. A well-known example is the defensive holding penalty called against the Philadelphia Eagles during their matchup with the Kansas City Chiefs in the 2023 Super Bowl[1].

On offense, infractions like holding or false starts can push a team back, making it more difficult to convert on downs and score. On defense, penalties such as pass interference or roughing the passer can extend the opponent's drive and increase their chances of scoring. This project investigates the impact of penalties on team performance metrics and game outcomes, while also examining which teams tend to receive more penalties. By identifying penalty trends and their correlation with key performance indicators, the analysis aims to explore whether there are consistent patterns that could support or challenge the idea that NFL games are rigged.

2. Data:

This project uses three primary sources of data: NFL Penalties' penalty data[2], offensive and defensive statistics from The Football Database[3], and outcome metrics from NFL standings[4]. I collected data from the 2015-2024 NFL seasons to create a comprehensive set.

---

1 www.forbes.com
2 www.nflpenalties.com
3 www.footballdb.com
4 www.nfl.com

2.1 Penalty Data

The first sets of data I collected was from NFLPenalties.com, a website that tracks penalties committed by each NFL team across different seasons. For this project, I focused on overall penalty data along with two specific penalty types: offensive holding and defensive pass interference. These particular penalties are highly subjective and rely heavily on human judgment, making them ideal for examining potential officiating biases or patterns.

From the overall penalty data, I collected the following columns: Season, Team, Penalty Count, Penalty Yards, Beneficiary Count, Beneficiary Yards, Net Count, and Net Yards. For offensive holding, the columns include: Season, Team, Holding Count, and Opponent Holding Count. Lastly, from the defensive pass interference data, I collected: Season, Team, PI Count, PI Yards, Opponent PI Count, and Opponent PI Yards.

2.2 Offensive and Defensive Statistics

The next sets of data were collected from FootballDB.com, a website that provides detailed NFL statistics by team and season. I gathered both offensive and defensive statistics to evaluate how penalties might relate to performance and overall team success. From the offensive stats page, I collected the following columns: Season, Team, Total Points, Rushing Yards, Passing Yards, and Total Yards. From the defensive stats page, I gathered: Season, Team, Total Points Allowed, Rushing Yards Allowed, Passing Yards Allowed, and Total Yards Allowed.

2.3 Outcome Metrics

To evaluate how penalties and performance metrics relate to team success, I also collected data on wins, losses, and playoff qualifications from the “Standings” page on NFL.com. This page provides season-level summaries for each team, including their overall record and whether they advanced to the postseason. From this page, I collected the following columns: Season, Team, Wins, Losses, and Playoffs. While scraping the standings data, I used an if/else statement to identify whether a team qualified for the playoffs. On NFL.com, teams that made the playoffs had a superscript “x” next to their name. If the script detected the “x”, the team was assigned a 1 in the Playoffs column and assigned a 0 otherwise.

2.4 Data Scraping and Cleaning Process

I wrote web scraping scripts using Selenium to automate the process of visiting these pages and extracting the necessary data. To collect data from the past 10 NFL seasons, I used a while loop that iterated through each season. For each iteration, the script dynamically constructed the appropriate URL, scraped the corresponding data, and printed the season year and URL to confirm successful retrieval before proceeding to the next year. Each of the six datasets were stored in separate DataFrames.

The data required minimal cleaning overall, but some standardization was necessary to ensure consistency across all datasets. Because several NFL teams have changed locations in the past decade (e.g., the Chargers, Rams, and Raiders), I used a mapping dictionary to update and align the team names across seasons. I also resolved punctuation inconsistencies for the New York teams (changing N.Y. Giants to NY Giants) for uniformity. Additionally, when scraping the standings data from NFL.com, the site listed teams by mascot name (e.g., “Chiefs” instead of “Kansas City”), so I used another mapping dictionary to convert these entries to cities or states, which was used in the rest of the project.

After collecting and cleaning the DataFrames, I merged them into a single master dataset for analysis. Each of the DataFrames contained a Season and Team column, which served as the common keys for merging. Using these two columns, I performed a series of joins to combine the data into one comprehensive set that was then downloaded as a CSV file named “nfl_penalties.csv” that was used for further analysis.

The following table is a data dictionary that thoroughly describes the data in the final set:

Table 1

Field	**Type**	**Source**	**Description**
Season	Numeric	NFLPenalties	Year of the season
Team	Text	NFLPenalties	Location of the team
Wins	Numeric	NFL.com	Number of regular season wins
Losses	Numeric	NFL.com	Number of regular season losses
Playoffs	Binary	NFL.com	1 if the team made the playoffs, 0 if not

Penalty Count	Numeric	NFLPenalties	Number of penalties
Penalty Yards	Numeric	NFLPenalties	Number of penalty yards
Beneficiary Count	Numeric	NFLPenalties	Number of opponent penalties
Beneficiary Yards	Numeric	NFLPenalties	Number of opponent penalty yards
Net Count	Numeric	NFLPenalties	Difference between beneficiary and penalty count
Net Yards	Numeric	NFLPenalties	Difference between beneficiary and penalty yards
Holding Count	Numeric	NFLPenalties	Number of offensive holding penalties
Opponent Holding Count	Numeric	NFLPenalties	Number of opponent offensive holding penalties
PI Count	Numeric	NFLPenalties	Number of pass interference penalties
PI Yards	Numeric	NFLPenalties	Number of pass interference yards
Opponent PI Count	Numeric	NFLPenalties	Number of opponent pass interference penalties
Opponent PI Yards	Numeric	NFLPenalties	Number of opponent pass interference yards
Total Points	Numeric	Football DB	Total points scored in the season
Rushing Yards	Numeric	Football DB	Total rushing yards in the season
Passing Yards	Numeric	Football DB	Total passing yards in the season
Total Yards	Numeric	Football DB	Total yards in the season
Total Points Allowed	Numeric	Football DB	Total opponent points scored in the season
Rushing Yards Allowed	Numeric	Football DB	Total opponent rushing yards in the season
Passing Yards Allowed	Numeric	Football DB	Total opponent passing yards in the season
Total Yards Allowed	Numeric	Football DB	Total opponent yards in the season

3. **Analysis:**

3.1 How do penalties affect team performance?

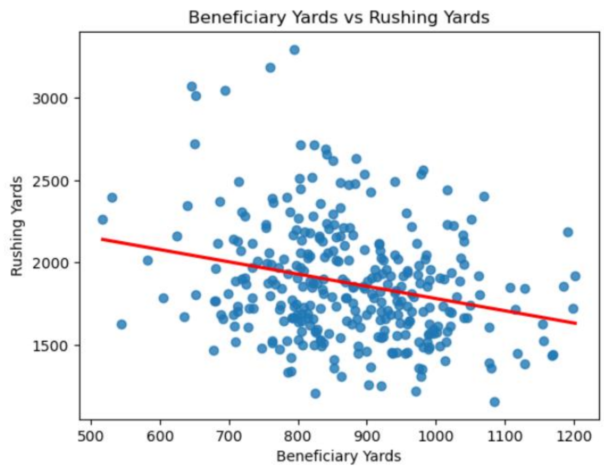
To begin my analysis, I examined whether teams that commit more penalties or receive more penalty yards tend to perform worse in terms of wins, playoff appearances, and overall performance metrics. I calculated Pearson correlation coefficients between key penalty variables and both offensive and defensive statistics.

Across the board, the relationships were weak. Teams with higher penalty counts or penalty yardage tended to have slightly worse outcomes in terms of Wins and Playoffs, but the correlations were minimal. Offensive performance metrics like Total Points and Total Yards also showed only weak or negligible associations with penalty counts. Beneficiary penalties showed slightly stronger positive correlations with passing-related metrics, particularly in cases like defensive pass interference, which can extend drives and result in large yardage gains. However, the overall impact on offensive production was still limited.

Defensive metrics showed similar results. There was still little to no meaningful relationship between the number of penalties a team committed and the points or yards they allowed. Even when isolating specific penalties, such as pass interference or opponent offensive holding, the data suggested that these infractions had little consistent influence on defensive success. Overall, while penalties are a visible and sometimes controversial part of the game, they do not appear to be strong predictors of team performance on either side of the ball.

One notable finding was the negative correlation between Beneficiary Yards and Rushing Yards with a coefficient of -0.258, suggesting that teams who benefit more from opponent penalties tend to gain fewer yards on the ground. The p-value was much smaller than 0.05, so the relationship is statistically significant, and visualized through a scatter plot in Figure 1. This relationship likely reflects the influence of pass interference calls. Since defensive pass interference can result in large gains and consequently more beneficiary yards, teams that gain more yardage this way might rely less on traditional rushing plays. This pattern suggests that teams benefitting more from these penalties may lean towards a more pass-heavy offensive strategy, leading to fewer rushing yards.

Figure 1

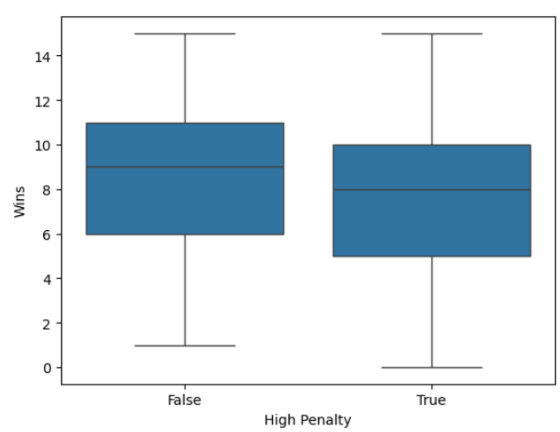


To further explore the impact of penalties on team performance, I grouped teams into two categories: those with above-median (High Penalty = True) and below-median (High Penalty = False) total penalty counts. I then compared the average values of key performance metrics across the two groups. The results are listed in the table below:

Table 2

High Penalty	Wins	Total Points	Total Points Allowed	Total Yards	Total Yards Allowed
True	7.881	370.125	374.819	5639.719	5632.613
False	8.456	376.456	371.763	5663.025	5670.131

Figure 2

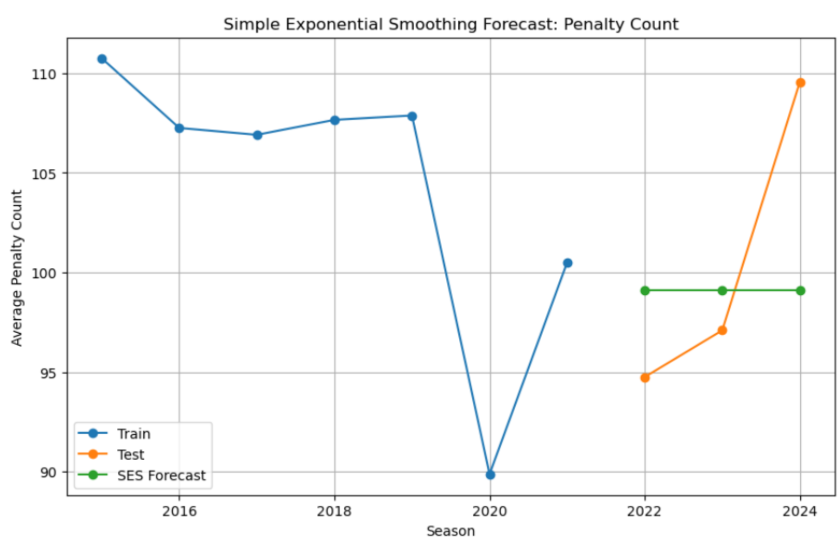


To test whether penalty data could predict playoff qualification, I built a logistic regression model using six penalty-related features: Penalty Count, Beneficiary Count, Holding Count, Opponent Holding Count, PI Count, and Opponent PI Count. The target variable was Playoffs. The model produced an accuracy of 64.1% and an ROC AUC score of 0.650, indicating modest predictive power. The confusion matrix showed that while the model could correctly classify a majority of non-playoff teams, with 31 true negatives, it struggled more with correctly identifying playoff teams, as indicated by the 17 false negatives. Overall, all of the results suggest that while penalties may have a small impact on performance, they are not strong standalone predictors of whether a team wins more games, qualifies for the playoffs, or has more offensive or defensive success.

3.2 Are there seasonal trends in penalties?

To explore whether there are seasonal trends in penalty counts, I used an exponential smoothing model to forecast future penalty totals. I tested two smoothing levels of 0.2 and 0.8. The model with a 0.8 smoothing level produced a lower mean absolute error of 5.604, compared to 7.283 for the model with 0.2, indicating it better captured recent trends in the data. I visualized the better model using a line graph, which highlighted the year-to-year fluctuations but no clear long-term upward or downward trend in penalties. These results suggest that while penalty counts vary slightly by season, there is currently no strong evidence of consistent seasonal trends.

Figure 3



I created two additional line graphs showing the average Holding Count and PI Count per team over the last 10 seasons. While there were no strong or consistent trends over time, one notable exception occurred in 2020, the season most affected by COVID-19. That year, the average Holding Count dropped to its lowest point, while PI Count spiked very high. The graphs are showcased in the following figures:

Figure 4

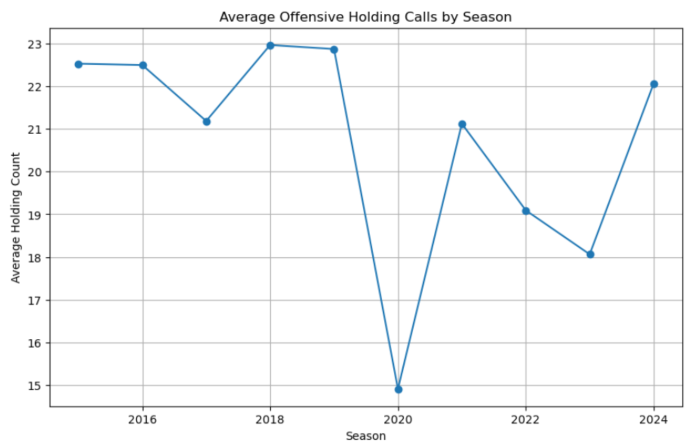
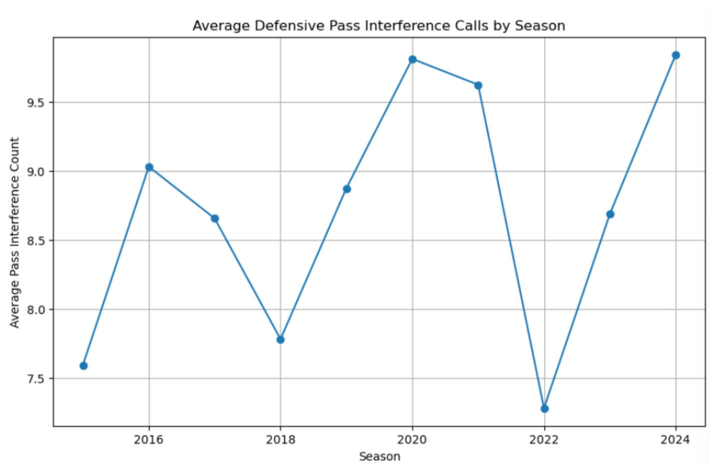


Figure 5



According to reporting from *ESPN*, this was not coincidental.[5] In 2020, at the direction of new league leadership, NFL officials dramatically reduced the enforcement of offensive holding. In the first three weeks of the 2020 season, 59% fewer holding calls were issued than in 2019 and defensive pass interference calls rose by 22%. Despite the rise in pass interference, the overall

[5] www.espn.com

number of penalties reached a historic low, as seen in Figure 3. These deliberate shifts in officiating priorities were likely aimed at promoting higher scoring and faster-paced games.

3.3 Are certain teams more or less likely to receive penalties?

To examine whether certain teams are more or less likely to receive penalties, I created two bar charts showing the total Penalty Count and Beneficiary Count for each NFL team.

Figure 6

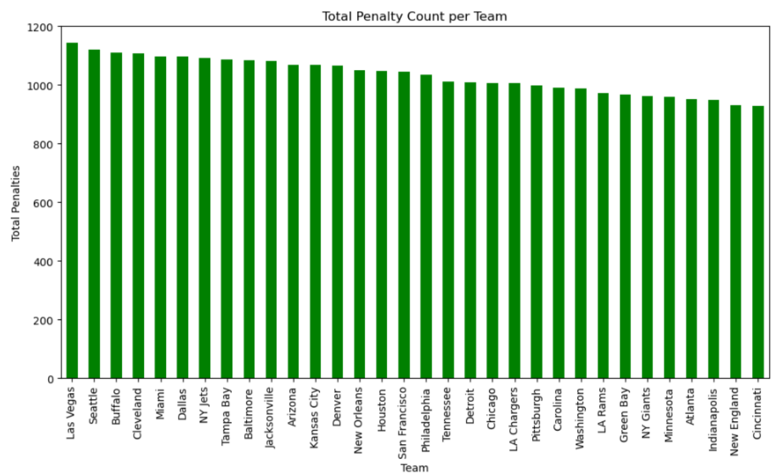
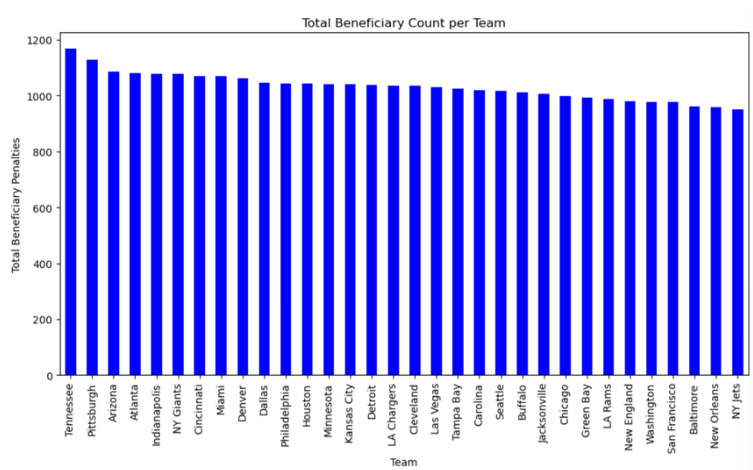


Figure 7

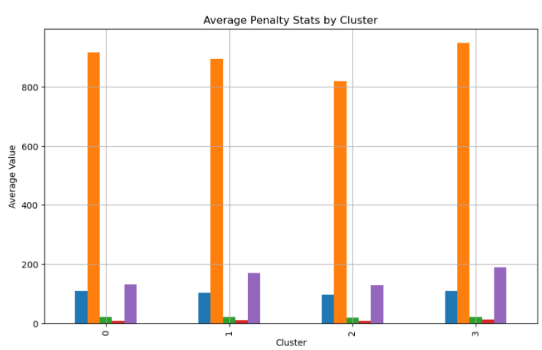


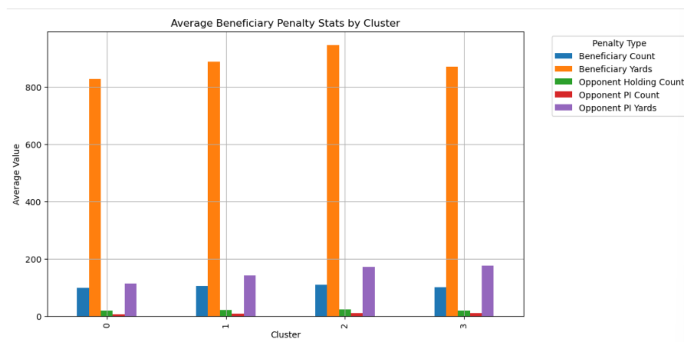
The bars were sorted from greatest to least to highlight the differences across teams. These visualizations revealed that the Las Vegas Raiders have received the most penalties over the last 10 seasons and the Cincinnati Bengals have received the least. The Tennessee Titans have drawn the most penalties from opponents and the New York Jets have drawn the least. Although this analysis does not confirm causality or bias, it does suggest that penalty distribution is not entirely uniform across teams, raising questions about play style, discipline, or even officiating tendencies.

To identify patterns in both penalty behavior and beneficiary trends, I applied K-Means clustering to two sets of features. The first clustering grouped teams based on penalties committed (Penalty Count, Penalty Yards, Holding Count, PI Count, PI Yards), while the second used penalties benefitted from (Beneficiary Count, Beneficiary Yards, Opponent Holding Count, Opponent PI Count, Opponent PI Yards). All features were scaled prior to clustering.

The results showed interesting contrasts across clusters. For example, teams like Kansas City, Baltimore, and Buffalo appeared in clusters with high penalty counts but low beneficiary metrics, suggesting they have been penalized more often than they benefitted. On the other hand, teams like Las Vegas, Indianapolis, and Tennessee were grouped in higher average beneficiary yardage and pass interference calls, indicating they consistently gained more from opponent penalties. The stacked bar charts compare each clusters' characteristics:

Figures 8 & 9





To visually compare which teams are most impacted by subjective penalties, I created two horizontal line charts. The first chart ranks teams by the total number of subjective penalties received, calculated as the sum of Holding Count and PI Count. The second chart shows which teams benefitted the most, using the combined total of Opponent Holding Count and Opponent PI Count. These charts highlighted noticeable variation across teams, suggesting that certain franchises may consistently draw more scrutiny, or more leniency, from officials when it comes to subjective penalties.

Figure 10

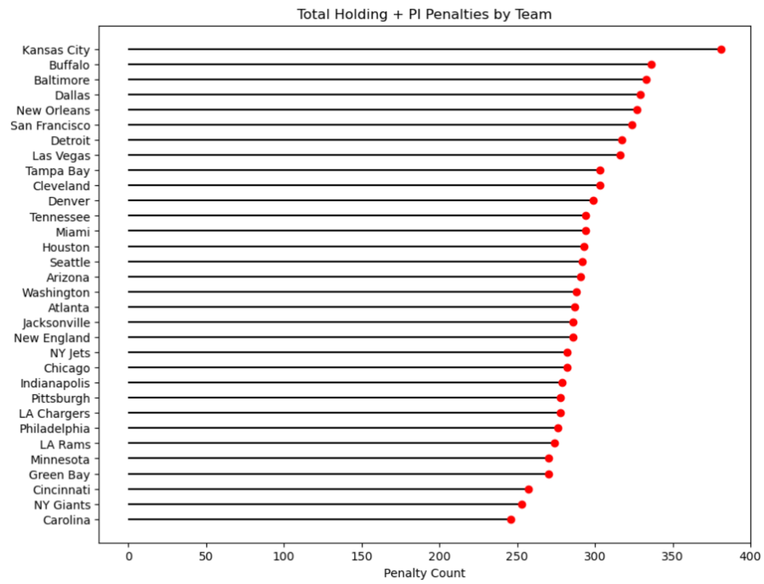
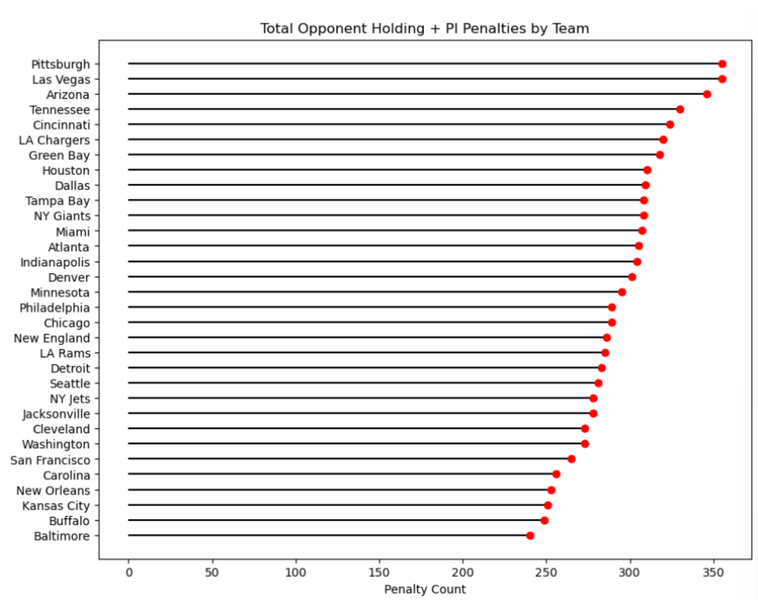


Figure 11



One surprising result from the line charts was that the Kansas City Chiefs received the highest number of subjective penalties and ranked 30th of benefitting from them. This finding stands in contrast to a common public perception that the Chiefs are favored by officials or benefit from biased officiating. While this doesn't rule out situational bias or missed calls, it does suggest that the Chiefs have actually been penalized more frequently by a wide margin (45 more penalties than the 2nd team, Buffalo Bills), contrasting the narrative that the game is rigged in their favor.

4. Conclusion

In this project, I explored how NFL penalties relate to team performance, seasonal patterns, and whether certain teams are more or less likely to be penalized. In summary, based on the three analysis questions I presented, I found the following results:

1. *How do penalties affect team performance?*

There were weak negative correlations between penalty metrics and team outcomes such as wins and playoff appearances. While penalties may slightly impact performance, they do not appear to be strong standalone predictors of team success. A logistic regression model using penalty data achieved only moderate accuracy in predicting playoff qualification.

2. *Are there seasonal trends in penalties?*

Exponential smoothing models showed minor year-to-year fluctuations in total penalties but no consistent upward or downward trend over time. However, when analyzing subjective penalties like offensive holding and defensive pass interference, the 2020 season stood out. Holding penalties sharply decline, while pass interference calls increased, likely due to officiating adjustments during the COVID-19 season aimed at encouraging more offense.

3. *Are certain teams more or less likely to receive penalties?*

Bar charts and clustering revealed that some teams are consistently penalized more than others, while others benefit more from opponent infractions. Interestingly, the Kansas City Chiefs, often cited in media narratives as benefitting from officiating, received the most subjective penalties over the last 10 seasons, challenging the idea that they are routinely favored.

This project has several limitations, including the lack of play-by-play context for when penalties occurred and no adjustment for play style or tempo across teams. Future work could include incorporating in-game context, examining penalties at the drive or play level, or analyzing referee crew tendencies. Additional insights could also be gained by combining this analysis with betting data or fan sentiment to further explore perceptions of bias versus statistical reality.