The National Football League: An Analysis of Thursday Night Football

Anna Ballou, Lizette Carpenter, Elaona Lemoto, Valerie Medina Cebreros, & Aoi Ogawa

Smith College

GitHub

Abstract

Thursday Night Football is an integral aspect of the National Football League's (NFL) brand. In recent years, however, viewers have questioned whether Thursday games are of the same caliber as non-Thursday games. This paper examines whether four key variables (travel distance, ELO win probabilities, expected points added, and run/pass completion rates) differ on Thursday and non-Thursday games. Several visualizations and permutation tests demonstrate limited evidence for a difference between game days. The NFL produces equitable games as there is no real difference between Thursday and non-Thursday games.

Introduction

The NFL, which has existed since 1920, has a mission to unite and inspire communities to enjoy the game through delivering the world's most exciting sports and entertainment experience. The League has four core values: Respect, integrity, responsibility to the team, and resiliency. While the NFL's mission is to deliver the "world's most exciting sports and entertainment experience," there have been many who question whether "Thursday Night Football" lives up to this integral part of the brand. Players, sports magazines and data analyst companies alike have all proclaimed that Thursday Night Football does not compare to the thrill of Sunday or Monday games. There are numerous theories that attempt to explain these lackluster Thursday games. On one hand, players, such as San Francisco cornerback Richard Sherman, cite the decrease in rest time (due to mid-week travel) as the culprit (*The Players'*

Tribune). Sports analysts, on the other hand, point to the number of uneven matchups that populate the Thursday game schedule as the reason for the uninspiring weekly performances (*FiveThirtyEight)*. Additionally, Bleacher Report boldly questions if Thursday Night Football could be hurting the NFL brand as a whole.

In this paper, we aim to investigate how the differences between Thursday and non-Thursday games manifest on game day. In doing so, we also intend to examine whether differences in team travel distances are potential causes of these game day manifestations. In an attempt to answer these questions, we will examine travel distance and game outcomes as a whole, as well as differences in play outcomes on Thursday vs. non-Thursday games. More specifically, we will investigate the following four variables: Travel distance, expected points added, run or pass play completion rates, and Elo win probabilities. This study will provide insight on the Thursday vs. non-Thursday phenomenon and will examine if travel distance is negatively impacting teams' performances overall. More broadly, this paper examines what (if any) future changes the NFL could implement in order to make football games as fair and entertaining as possible. We will examine whether the NFL may need to change how they schedule Thursday games (in taking into consideration how much teams travel) and if Thursday games are causing decreased player and team performance.

Data Description

The majority of the data we will use in this study is owned and maintained by the NFL. We will also use variables from other sources (FiveThirtyEight and a separate stadium dataset). The two datasets from the NFL are "Play by Play" and "Games." Play by Play represents different elements of games at the play level. Some variable examples include: the score before and after a play, timouts, possession team, and type of play (run/pass). Key variables from Play by Play that we will be using are the following:

1. Play_description (a characteristic): a description of individual play which includes important players and play-specific detail.
2. pass_result (a factor): a description of whether the play was completed, incomplete, intercepted, sacked , and scrambled.
3. play_result (an integer): depicts the yardage that was either gained or lost after the play was executed.
4. home and visitor_score_after_play (an integer): consist of the score variable that changes within a game.
5. Quarter (a factor): Quarter during which the play occured. Football typically has 4 quarters but may have 5 for overtime.
6. Ep (an integer): denotes the expected points added as a result of a play.

The Games data provides higher-level, overarching information about each game. For example, information about weather, kick-off time, and game day. From Games, key variables we will use are the following:

1. Season (a factor): represents the season when the game took place. Denoted by year (2007, 2008, … 2019).
2. Game_day (a factor): consists of the day of the week in which the game took place.
3. Game_site (a factor): consists of the site of the game. It can represent either by the city name or the name of the team.

Additionally, we will use an external dataset with information regarding the team's stadiums. This “stadium dataset” was created from the coordinates for each stadium. In this dataset, we see the following variables:

1. Season (a factor): represents the season where the game took place. Denoted by year (2007, 2008, … 2019).
2. Home_Team (a factor): represents the home team in the game.
3. Visit_Team (a factor): represents the visit team in the game.
4. Visit_lat (a numeric): represents the latitude of the visit team’s stadium location.
5. Visit_long (a numeric): represents the longitude of the visit team’s stadium location.
6. Home_lat (a numeric): represents the latitude of home team’s stadium location.
7. Home_long (a numeric): represents the longitude of home team’s stadium location.

Lastly, we will use an external data set pertaining to the Elo ratings to obtain the differences in Thursday vs. non-Thursday games and distance traveled. From the nfl_elo we will use the following variables:

1. date (date): date of game
2. Team1 (factor): Abbreviation for home team
3. Team2 (factor): Abbreviation for away team
4. Elo1_pre (a numeric): Home team's Elo rating before the game
5. Elo2_pre (a numeric): Away team's Elo rating before the game
6. Elo_prob1 (a numeric): Home team's probability of winning according to the Elo rating
7. Elo_prob 2 (a numeric): Away team's probability of winning according to the Elo rating
8. Season (a factor): Year of season

Methods

Once we obtained the data from the NFL, we examined the variables to gain an understanding of potential topics. We researched football terminology, as some of the unique phrases and variable names within the dataset were unknown to us. In order to develop our own research question, we examined what research questions were currently being asked by NFL sports statisticians. After gaining inspiration through current work, we conducted our own exploratory data analysis on variables we deemed interesting and important.

Our next step was to combine our different exploratory data analyses into one overarching research question. We identified four variables of interest: travel distance, expected points added, run or pass play completion, and Elo win probabilities. We examined each of these variables for differences on Thursday and non-Thursday games through both visualizations and formal permutation tests. In each test, we found an observed difference between the two game day types and performed numerous iterations (at least n = 1,000). The specifics of each test and visualization for each unique variable are outlined in the results section.

Results

Travel Distance

One key variable we examined for differences in Thursday and non-Thursday games was travel distance. Since Thursday games are mid-week, we aimed to investigate if teams were traveling less or more miles. Significant data exploration was required for travel distance. The visualizations below are a combination of general travel distance exploration (not related to game day) and Thursday vs. non-Thursday analyses.

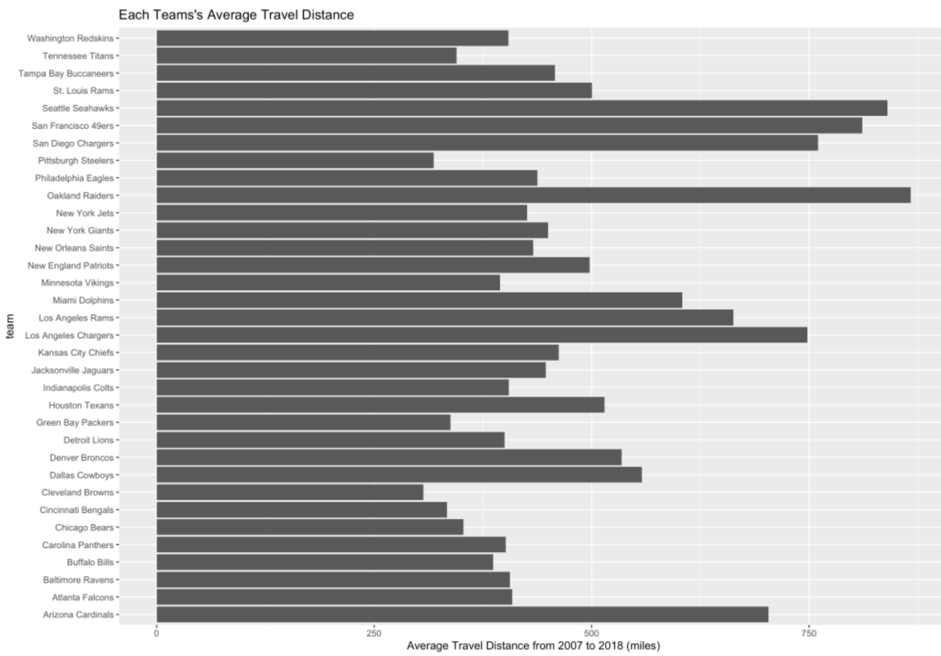

Figure 1:Team's average travel distances

Figure 1 shows the average distance that NFL teams traveled during the 2007 to 2018 football seasons. The team (out of the 32 NFL teams) that traveled the most during this time period was the Oakland Raiders (Oakland, CA) with an average of 826.72 miles per year. Noticeably, the team that traveled the least was the Pittsburgh Steelers (Pittsburg, PA) with an average of 303.46 miles per year. Looking at both teams in particular we found that the difference in their travel distance was 523.26 miles.

To analyze whether teams traveled further for Thursday games, a permutation test was performed. We hypothesized that there would be no difference in mean travel distance between

Thursday games and non-Thursday games. The observed difference in mean travel distance between Thursday games and non-Thursday games was 78.21. The distribution of travel distance is shown in figure 2 and shows that teams seem to travel further for Thursday games. For the permutation test there were 5000 samples generated and each sample mean difference was calculated. The result showed that there was no significant difference in mean travel distance between Thursday and non-Thursday games (p-value = 0.094, Appendix).

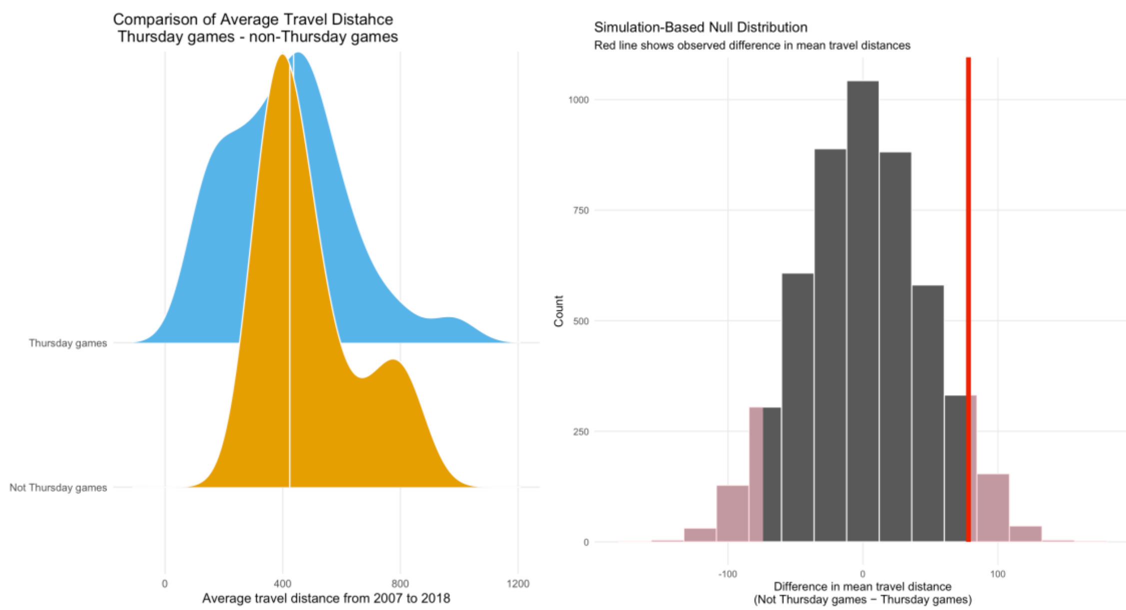

Figure 2: Distribution of average travel distance on Thursday and non-Thursday games (right shows the Distribution of travel distance and the left shows the null distribution)

Elo Rankings

Data from FiveThirtyEight was incorporated in our study to evaluate elo rankings. Elo ratings is a simple system that judges teams or players based on head-to-head results. FiveThirtyEight has their own unique system that follows Elo ratings where they adjust for various variables. The FiveThirtyEight NFL Elo rankings can be used to forecast the outcome of every game in every season. For example, we can track how the winning probability for

Pittsburg changes throughout a season if they are the home or visiting team (Appendix, Figures A2, A3). Considering how prevalent elo rankings are in the NFL, we wanted to use this as a variable in our study. We focused on how the elo winning probability for the home and away team before a game compares based on Thursday and non-Thursday games and distances traveled. We hypothesized that based on the potential amount of training and resting time loss by playing on Thursday, then the winning probability would be lower for Thursday games. Concerning traveling for a game, we hypothesized that the winning probability for the away team would be lower for Thursday games.

We compared the distribution of mean winning probability for home teams on Thursday and non-Thursday games between 2007 and 2019 (Figure 3). From the distribution, we do not see a significant difference of mean winning probability between Thursday and non-Thursday games. However, we proceeded with a permutation test for a difference in mean winning probability for home teams on Thursday and non-Thursday games. The test showed an observed difference in winning probability for home teams of 1.64 elo ranking (non-Thursday - Thursday). Depicting that home teams perform on average slightly better on non-Thursday games than on Thursday games. Yet, the p-value for this test is not statistically significant with a value of 0.926. We proceeded in a similar matter for comparing the distribution of mean winning probability for away teams on Thursday and non-Thursday games using a permutation test.

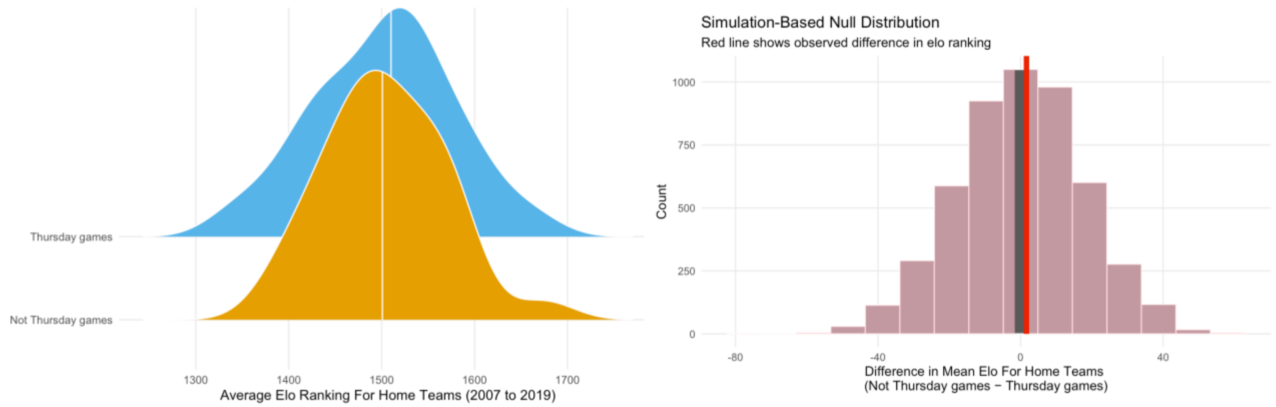

Figure 3: Right shows the distribution of average elo rankings for Home Teams on Thursday and non-Thursday games. Left shows the null distribution.

The permutation test conducted for the difference in mean winning probability for away teams on Thursday and non-Thursday games showed a slightly greater difference than our permutation test for home teams (Figure 4). This test showed an observed difference in winning probability for away teams of -17.56 (non-Thursday - Thursday) meaning away teams perform on average better on Thursday games than on non-Thursday games, which is opposite to what we expected. Since we were also interested in how distance travelled may have affected the winning probability for a team we decided to create a distribution of the difference in winning probability on close, far and very far games (Appendix, figure A4).

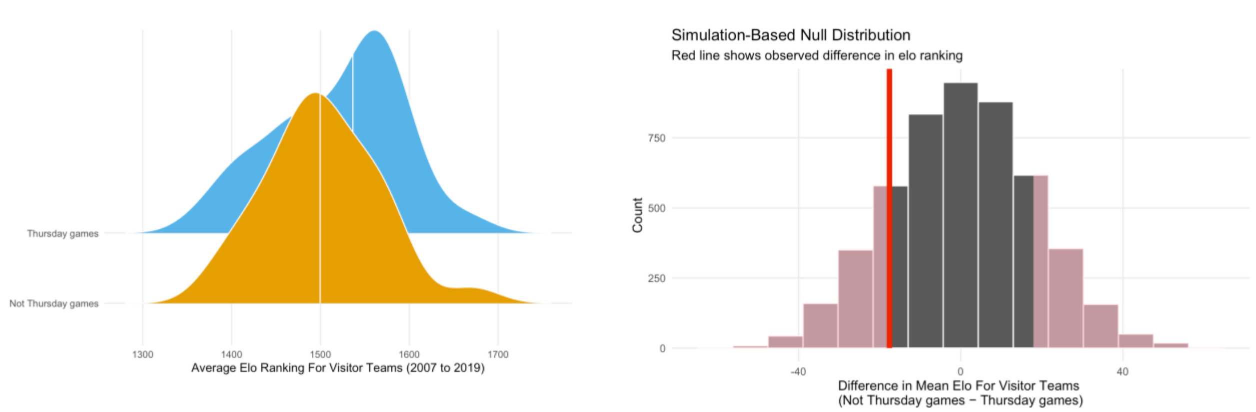

Figure 4: Right shows the distribution of average elo rankings for Away Teams on Thursday and non-Thursday games. Left shows the null distribution.

We created a data visualization that maps the distances travelled by a team in a season (Figure 5). This visualization also shows the winning probability for the home and away teams.

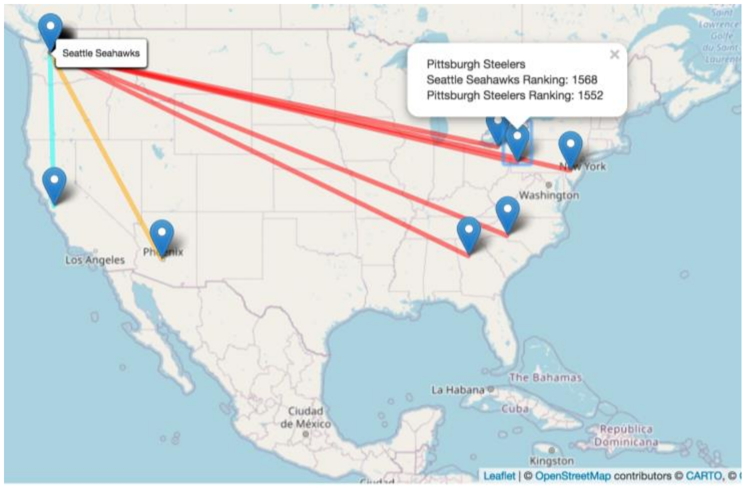

Figure 5: Distances travelled by Seattle Seahawks for away games in 2019 with Elo rankings information.

EPA

Another potential variable we examined was expected points added (EPA). While the NFL dataset provided a variable for expected points (EP), we updated this variable to incorporate cumulative increases and decreases in expected points from one play to the next (i.e. EPA). After each play, EPA depicts whether a team is closer or farther from earning points. For example, a completed 80-yard pass would have a high EPA since the chances of scoring a touchdown increased dramatically from the previous play. Contrarily, a sack (when the quarterback is tackled before releasing the ball) would result in a considerable drop in EPA as the likelihood of scoring a touchdown is decreased from the prior play. Some football viewers believe that Thursday games are less prolific and have fewer exciting, risky, game changing plays. Hence, we hypothesized that EPA would be lower on Thursday games.

In comparing the distribution of mean EPA on Thursdays and non-Thursday games (Appendix, Figure A1), no clear difference was evident. This initially suggested that there was no significant difference between EPA on Thursday and non-Thursday games. To test this claim, we conducted a formal permutation test for a difference in mean EPA on Thursday and non-Thursday games. The test found an observed difference (non-Thursday minus Thursday) of -0.015. This suggested that Thursday EPA was larger than non-Thursday EPA, which was contrary to what we expected. The p-value of this test, however, was 0.935 (not significant at the α = 0.05 level). The results of the permutation test are displayed in figure 6.

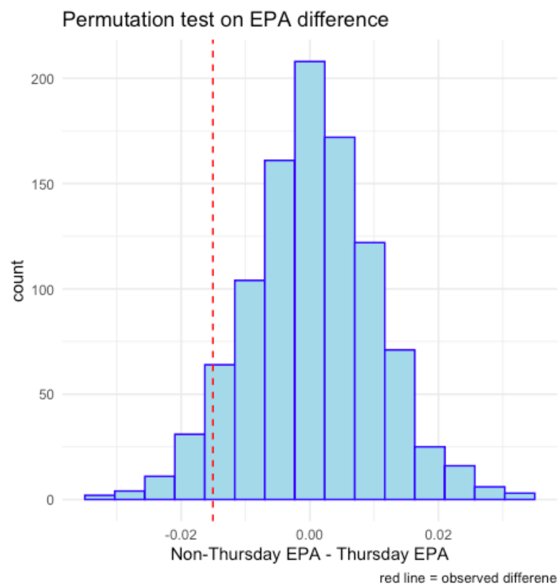

Figure 6: Results of a permutation test for difference in mean EPA. The observed difference is in red. There were 1,000 iterations in this test.

We hypothesized that one potential reason for the non-significant results of this test was variation on a team-by-team basis. For example, when comparing EPA on Thursday and non-Thursday for certain teams, we noticed that high-powered teams had high EPAs for all game days (and vice versa). The EPAs for these teams could be skewing the mean EPAs examined in the permutation test. Thus, we compared EPA on Thursday and non-Thursday games for teams with both winning and losing seasons. We hypothesized that these teams might be affected by differences in Thursday and non-Thursday games more strongly and that trends in EPA might be more obvious. An analysis of this idea showed a clear difference in EPA on Thursday and non-Thursday games for these teams (figure 7). Additionally, most teams saw lower average EPA on Thursday games. This phenomenon matched our initial hypothesis that Thursday games would have lower EPAs. The severity of the difference, however, varied sizably between teams.

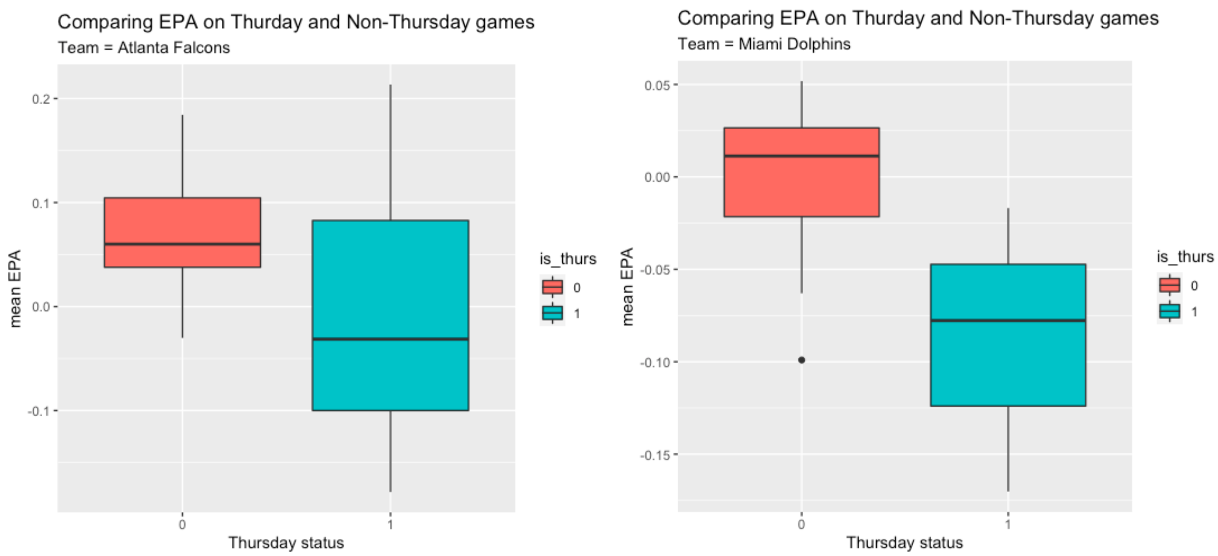

Figure 7: Comparing mean EPA on Thursday and non-Thursday for teams with both winning and losing seasons during the time period.

Completion and Positive Run Percentage

One of our last interests in this project was looking at team performances and seeing if there were any clear differences in how teams approached different facets of game conditions (ie., playing Thursday games, traveling to different cities, states, or countries). In the beginning of the project, we hypothesized that since Thursday games appear to be the least exciting game day, teams would be more inclined to execute riskier plays. When looking at Thursday games, we might see a higher average of risky plays and lower average for non-Thursday games. For distance games, we wanted to know if there was any evidence of travel exhaustion on team performance; thus a higher average of both statistics for close games. Overall, our goal for this part of our project looks to ensure that the NFL carries out a system that allows an equitable playing field in that team performance stays relatively the same regardless of the distance traveled or the day played.

To measure team performance, we decided to look at two different measurements: completion and positive run percentage. Completion is the rate of completed passes and is calculated by taking the proportion of passes that were caught by the possession team divided by the number of completed passes, intercepted passes and incomplete passes. As for positive run percentage, we took the number of plays where the QB made a run play which made positive yards divided by the total number of positive and negative yards gained from a running play. We chose these two statistics since we thought that showing offensive play might be indicative of overall team performance.

Now in creating these proportions, we wanted to test if there was any difference in team performance by Thursday and Non-Thursday games and by travel distance. Before running our tests, we decided to run preliminary exploratory data analysis. If you notice with Figure 8 below, we see that the average completion rate among Thursdays and non-Thursdays is very small.

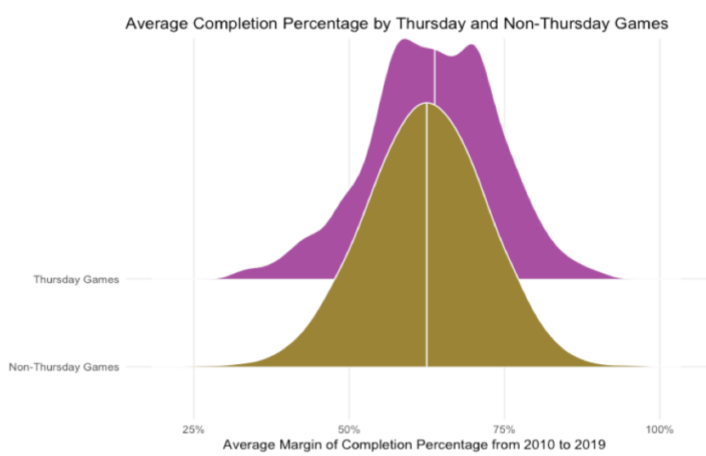

Figure 8: Completion rate on Thursday and non-Thursday games

We also see that the distributions are relatively the same as well. We found that the average completion rate is higher on Thursday games, which gets at our hypothesis earlier. Since

Thursday games rarely showcase important matches for teams, there might be more risk in teams throwing the ball more often. After running our permutation tests, we found that there was no statistical significance with a p-value of .256. Then looking at positive run percentage, we saw that there was a slight difference where Non-thursday games seemed to have a higher average in positive run percentages, again what we hypothesized prior to our tests. However, after running our permutation tests, we saw that the associated p-value for the difference in means was .182 so again not statistically significant. What we see here is that there isn't a significant difference in team performance between game days.

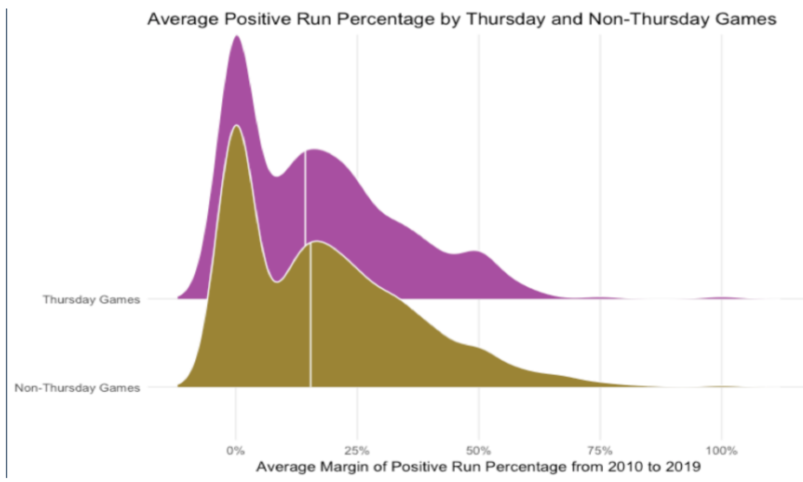

Figure 9: Average Margin of Run Percentage

Lastly, looking at team performance as a function of travel distance, we found some interesting results. When looking at completion percentage, we saw that there was relatively no difference in means among the three categories of distance. We then ran an F-test on the relationship between the categories of distance on completion and we had a very high p-value of .446 thus telling us that completion is not significantly different among travel times. When looking at positive run percentage as a function of our three categories of travel distance, we saw a slightly higher average in close games, which coincided with our hypothesis of better performance. However, after running our permutation tests with our F-statistics, we found a p-value of .698. With this p-value, it's more likely that distance has no effect on positive run. Like mentioned earlier, due to the small sample sizes, there's only so much our tests can do. Overall, we found no significant effect on team performance, but further research and more comprehensive data could inform us otherwise.

Discussion: Challenges, limitations and Future Directions

One major challenge we faced was understanding football terms, since none of us had experience in football. There were a variety of football terms that we did not understand, but we were able to find definitions through our contacts at the NFL. Lastly, we struggled with narrowing down our research question. We had numerous ideas and potential research pathways, making it difficult to find a concise yet intriguing objective. Fortunately, we were able to overcome this hurdle by formulating an overarching question that combined many of our research interests and ideas.

One major limitation of our analyses was that the NFL data was missing a large amount of information for the 2007-2010 seasons. Because of this, we questioned if that missing data would have had an impact on our results and findings. Another limitation we considered was that the NFL institutes major rule changes every year. This makes it difficult to generalize our results to future years. This is especially true for the coming 2020 season, as COVID-19 will be a huge factor in scheduling future games. We were also concerned about the sample size of our data. On average, teams play one or two Thursday games a season compared to over twelve non-Thursday games. These drastic differences in sample size may have impacted our results as we were comparing largely different sized groups. We attempted to ameliorate this difference through running permutation and bootstrap style tests, but the sheer small size of the Thursday group may have still impacted our findings.

In the future, the NFL can continue work on the Thursday phenomenon through examining if trends we found continue in future seasons. Season-to-season trends of the variables we considered would only heighten the power of our findings. Additionally, the NFL could examine other potential variables that might be impacting Thursday vs. non-Thursday differences. We only examined four factors, but the dataset included upwards of 50 possible predictors. One such example includes if a player was injured or traded to a different team. Additionally, we suggest examining differences between West and East coast teams. We noticed that, on average, West Coast teams tend to travel farther for games. We are curious if other differences exist between West and East coast teams and if these differences might impact Thursday play.

Lastly, in terms of the ethics of this project, we were concerned about transparency. The majority of the data that we used in our analyses is owned by the NFL. Throughout our analyses, we were conscious of this fact as the NFL could potentially manipulate the data. While this was seemingly unlikely, it was still a concern of ours. This led us to specifically question the missing data from seasons 2007-2010. We were never told what happened to that data or why it was not included in our datasets.

<u>Conclusion</u>

We found that, for the four variables considered, no real statistical difference existed between Thursday and non-Thursday games. All of our tests resulted in non-significant p-values (at $\alpha = 0.05$) and showed a marginal difference between Thursday and non-Thursday games. Thus, we conclude that, in considering these four variables, the NFL produces equitable games with no significant differences between Thursday and non-Thursday game days. Additionally, travel time does not seem to have an impact on teams' performances on gameday.

It is important to note, however, that because of the small sample size, some marginal differences between Thursday and non-Thursday games could potentially be magnified if a larger Thursday sample was possible. Thus, while we did not meet a specific p-value threshold in our analyses, it is still important to note that we did observe some potential differences. Namely, some of our results showed that teams had lower levels of varying performance metrics for Thursday games. While our overall conclusion comes from only examining four potential factors, we are confident that we chose the four most important variables that may produce a potential difference. We do, however, encourage future work in examining the Thursday Night Football conundrum.

Appendix

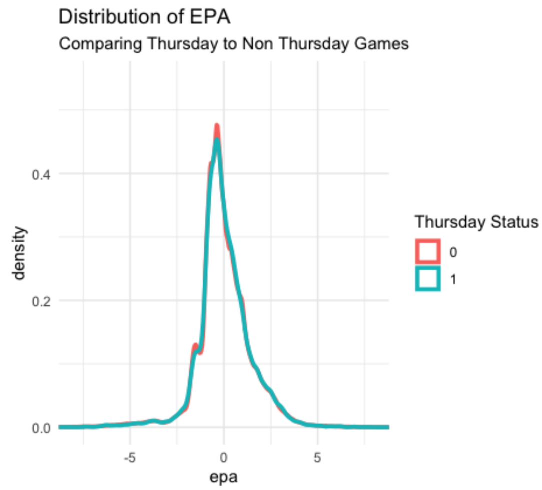

Figure A1: Distribution of mean EPA on Thursday and non-Thursday games. Suggests no significant difference between games

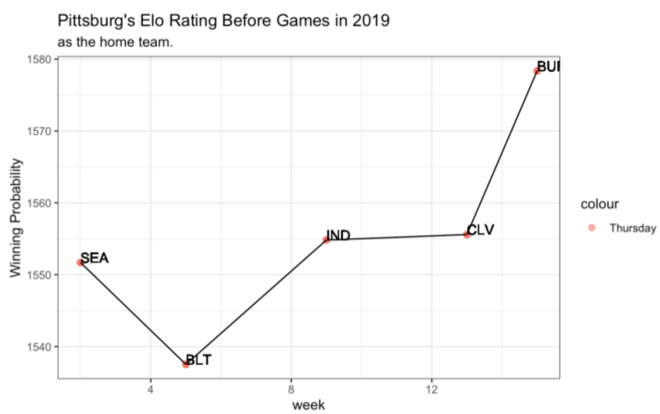

Figure A2 : Pittsburgh's Elo Winning Probability as Home Team in 2019

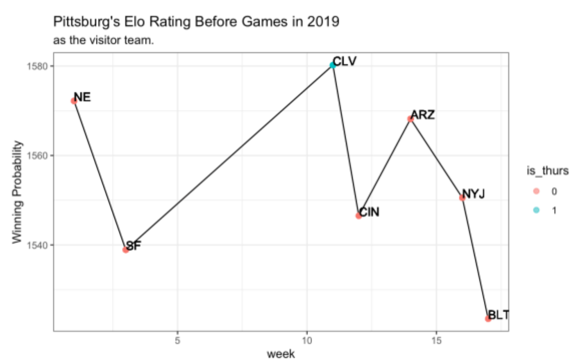

Figure A3 : Pittsburgh's Elo Winning Probability as Away Team in 2019
Figure A4: Distribution of average elo rankings for away teams on distances travelled between 2007 and 2019.

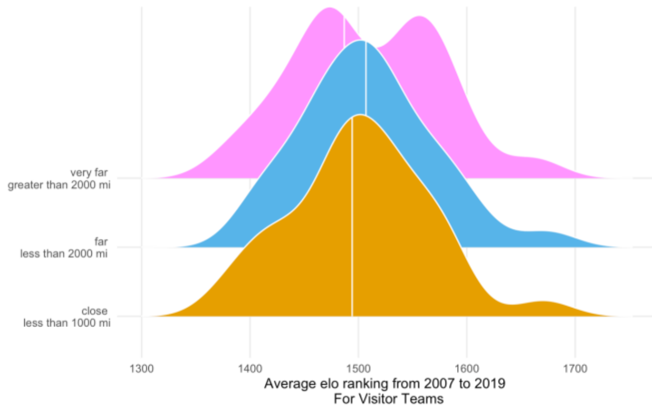

Figure A4: Comparison of average elo rankings by travel distance (categorized by far, very far, and close).

References

- ELO ratings: https://data.fivethirtyeight.com/#nfl-elo
- Sherman , R. (2016, December 14). Why I Hate Thursday Night Football. Retrieved 2020, from https://www.theplayerstribune.com/en-us/articles/richard-sherman-seahawks-thursday-night-football
- 2019 NFL Predictions: NateSilver538 - https://projects.fivethirtyeight.com/2019-nfl-predictions/
- Stadium Coordinates (One Example: FedExField.(2020, April 7). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/FedExField). Found for all NFL teams.

- Finnis , A. (2018, October 5). NFL glossary: All the terms you need to understand the game, from audibles to zone reads Retrieved 2020, from https://www.telegraph.co.uk/american-football/0/nfl-glossary-terms-need-understand-game-audibles-zone-reads/