Introducing: PlayCallR

A Tool to Aid Key Gametime Decisions for NFL Coaches

Final Report

Prepared for Dr. Donald Wedding, CEO
March 1, 2022

GameBreakR Analytics

Kristen Allen
Shruthi Harve Iyengar
Max Pauly
Michael Venit
Alex Zhou

Play Call Summary

Q1: Problem and solution

Picture this…

You're the head coach on an NFL Sunday. It's 4th down and 2 yards to go on the opponent's 35. You're down 3 points with 4:00 left in the fourth quarter. Your defense has held the opponent to 10 points and your kicker is facing a light wind. What is the play call? Do you go for it, attempt the 52-yard field goal, or punt?

Introducing: PlayCallR

PlayCallR, a cutting-edge app from **GameBreakR Analytics**, recommends a play call based on game context, complete with likelihood of success and the impact on win probability.

THE WHAT

The PlayCallR app provides objective recommendations on critical NFL play calling decisions, adding a dynamic tool to each team's arsenal.

THE WHO

The NFL should offer the PlayCallR app to all 32 NFL teams.

THE WHY

More effective play calling decisions leads to more competitive NFL games, leading to increased viewership, fan engagement, and revenue.

- Equitable access to the PlayCallR analytic tool levels the playing field for NFL teams with varying degrees of analytic capability
- Access to one data tool may lead to job growth for teams investing in data scientists and analysts
- An objective tool assists the coach but does not replace the coach's experience and expertise as it affects play calling
- The app version of the PlayCallR tool is environmentally conscious and provides a user-friendly interface
- The PlayCallR tool can be built in any data visualization platform, including Microsoft's Power BI. The NFL can leverage its existing relationship with Microsoft to negotiate low maintenance costs

Situation Analysis

The NFL currently has rules that prohibit coaches from using technology on the sidelines, with the exception of game film provided in partnership with Microsoft. There are currently no standardized analytics models available to all teams. But the implementation of PlayCallR for all 32 teams will level the analytics playing field in a situation where some teams have robust analytics teams and others do not. GameBreakR Analytics will work with the league office and the chairman of the NFL Competition Committee to understand our options and see if we can provide the app on the sideline or just the resulting paper-based decision grid.

PlayCallR would provide coaches with information they did not previously have when making decisions during the game. This would lead to more competitive games, which in turn would increase TV and ad revenue. Currently, the NFL brings in close over $10B annually, with media rights and sponsorships accounting for $1.62B annually. Increasing the competitiveness of games would increase viewership and the media rights contracts, as the major networks want a competitive product to showcase.

HOW IT WORKS

PlayCallR is a tool that makes recommendations for two critical play calling situations:

Fourth Down:
Punt, Field Goal or "Go For It"

Points After Touchdown:
Extra Point Kick vs. 2-Point Conversion

The fourth down model evaluates the conversion rates of going for it on fourth down, punting the ball, and kicking a field goal and provides an optimal recommendation based on the increase in Win Probability (Win Probability Added) from the decision. The model takes into account the conversion rate of the fourth down and distance, the conversion rate of a field goal number of yards to the goal line, game score, time remaining, and number of yards to a first down.

The PAT (Points After Touchdown) model takes into account the time left in the game, the conversion rate of going for 2 points, the conversion rate of kicking the extra point, and what the score is and recommends the optimal decision based upon the Win Probability Added.

Q2: Data Acquisition

About the Data

The data that we are using is NFL play by play data for the regular season from 2015 to 2021, obtained using the nflfastR[1] library in rStudio. The data is maintained on a consistent basis throughout the season, updating as each game is played. This allows for the most recent data to be available for analysis and for our model.

The dataset includes:

- 340,243 individual plays from 1,888 games (every play for the last 7 seasons)
- 372 variables on each play

We were able to take a unique approach to the feature selection process, as one of our team members has been studying and working with this data for a few years. We leaned on Michael Venit's domain expertise to determine which independent variables would impact our dependent variable, Win Probability Added. This helped us eliminate the categorical variables that we did not need, and it also helped us identify the numerical variables that would have the biggest impact on the predicted value. The Exploratory Data Analysis in the next section also informed feature selection. We also evaluated Principal Component Analysis for variable selection and found little increase in accuracy, so we kept the domain expert's selected variables.

Some variables were of interest to us, such as temperature, time of day, stadium, wind and wind direction. We hypothesized that each of these could impact decision making, especially wind and weather since we are providing a recommendation on kicking a field goal or not. However, not all games reported this data consistently, so we needed to omit these variables from the model.

We also removed specific cases from the model that would impact its accuracy, such as:

- Fourth down plays that resulted in a penalty (5.4%) or a quarterback kneel down (.07%)
- Plays that begin inside the offensive team's 35-yard line (65 yards or more from the endzone). Since these situations would almost always result in a punt recommendation, the PlayCallR app is not needed for these plays. This was confirmed with an industry expert in an analytics role with an NFL team.

[1] https://www.nflfastr.com/

Exploratory Data Analysis

Fourth Down Data

The first two graphs in **Appendix A** show fourth down conversions and failures by play type and by quarter.

- Conversions far outweigh failures in the first 3 quarters, perhaps due to more conservative play calling earlier in the game.
- In the fourth quarter, attempts go up significantly and failures increase.
- Additionally, pass is the most common play type for the fourth quarter attempts, while run and pass are fairly even during the first 3 quarters.

The next three graphs in **Appendix A** show fourth down successes, failures and attempts by play type.

- Overall, runs are more successful than passes, which probably is more about the situation in the game than run plays being the best choice.
- The third graph illustrates that only Baltimore and Philadelphia are successful on 4th down conversion attempts more than half the time.

PAT (Points after Touchdown) Data

The three graphs in **Appendix B** show us general trends in play calling for two-point conversion attempts.

- Pass plays are attempted more frequently, but rushing plays have a higher success rate.
- The total number of two-point conversion attempts is relatively low, occurring on 8% of points after touchdown plays.
- Some teams are more likely to attempt the two-point conversion than others.

Q3: Analytics Model

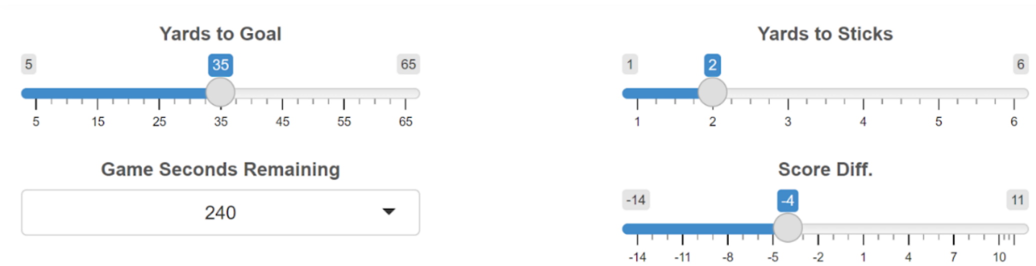
The outcome of the models is PlayCallR, an interactive app shown [here](). The app consists of recommendation engines for Fourth-Down and Points After Touchdown situations. Overall, the implementation of PlayCallR offers NFL teams an additional analytic tool that is objective and levels the playing field. The use of PlayCallR as one factor in the coach's decision-making process can lead to more aggressive play calling and increased competitiveness.

Fourth-Down Play Calling Recommendations

The aggregated model allows a coach to input these four variables:

Time Remaining	Distance to First Down	Distance to End Zone	Point Differential

The model then returns a recommendation to the coach as to whether they should punt, go for it, or kick a field goal. As an example, this team is down by 4 points with 4 minutes remaining in the game. They are on the opponents 35-yard line and have 2 yards to gain for the first down. PlayCallR is recommending that the coach go for it, rather than attempt a 52-yard field goal.



Go For It (+2.8)

4th & 2, 35 yards to goal, Down -4, 240 seconds left

Label	WPA if Success	WPA if Failure	Exp. Conversion Rate	Expected WPA
Go For It	14	-11.3	57.5	**3.3**
Kick a FG	10	-14.5	61.2	**0.5**
Punt	NA	NA	NA	NA

Model Construction

The fourth down recommendation engine in PlayCallR is an aggregate of 6 different models using the variables listed in **Appendix C**:

- A logistic regression model to predict fourth down conversion rate
- A logistic regression model to predict field goal conversion rate
- Two random forest models to predict win probability added (WPA)--both with successful and unsuccessful fourth down conversion attempts
- Two random forest models to predict win probability added (WPA)--both with field goals made and missed

All 6 models are then used to calculate two figures, the "go for it" WPA (win probability added) and the field goal attempt WPA (win probability added). In the example shown above, this is calculated as follows:

Go For It

$$0.575\left(P_{success}\right) * 14(WPA) + 0.425\left(P_{failure}\right) * \quad - 11.3(WPA) = 3.3\,WPA$$

Field Goal

$$0.612\left(P_{success}\right) * 10(WPA) + 0.388\left(P_{failure}\right) * \quad - 14.5(WPA) = 0.5\,WPA$$

The difference between the outputs is **2.8 higher for "go for it"**, which feeds the recommendation in PlayCallR.

Note that PlayCallR will return a recommendation of "toss up" if the WPA (win probability added) differential is under 2 percentage points, leaving this play call to the coach's discretion.

Model Accuracy

Prediction Accuracy

Fourth-Down	Field Goal
61.1%	84.6%

R^2 Values[2]

	Fourth-Down	Field Goal
Success	0.6303	0.7490
Failure	0.6344	0.7453

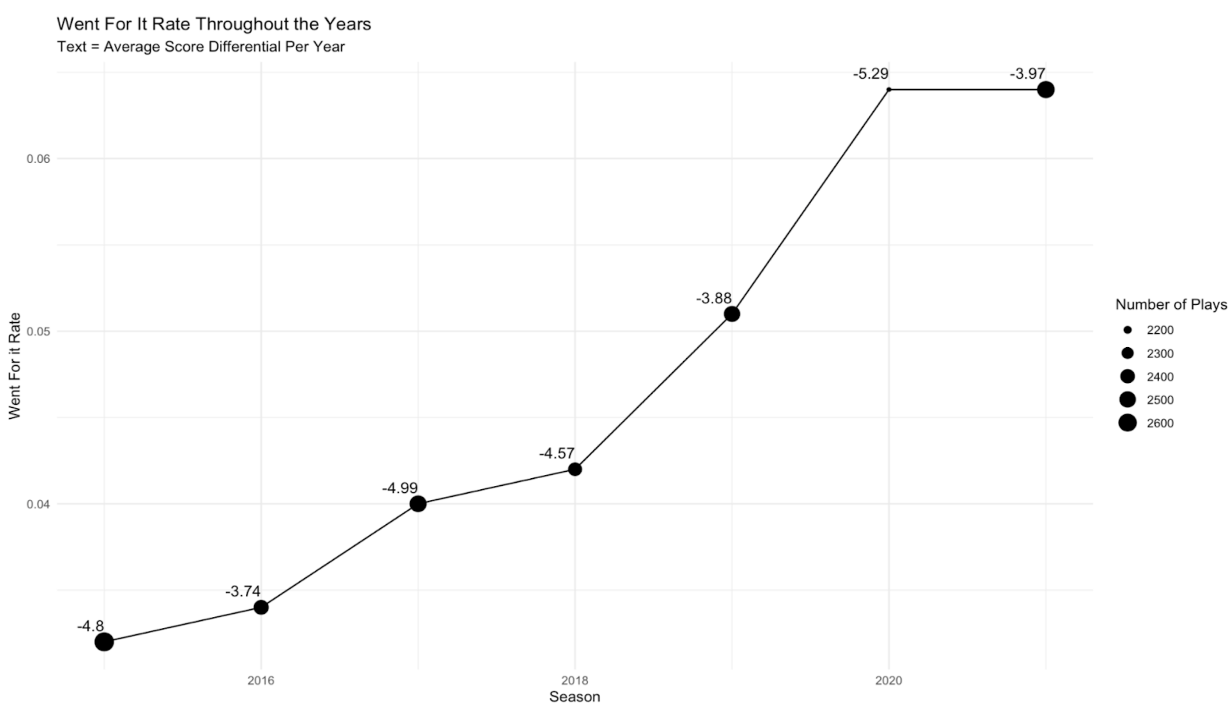
Overall, all 6 models provide fairly strong accuracy measures. The accuracies of the fourth-down conversion models are slightly lower than those of the field goals perhaps due to wild card factors, such as star talent or home/away status, which are not captured in the

[2] R^2 value is the percentage of the variability in the response that is explained by the model. The R^2 values shown here were generated by our 4 random forest models

models. However, it is still better than random choice, which has an assumed 50% accuracy rate.

Insights

- Not surprisingly, the distance of the field goal is the most important metric in determining success.
- Similarly, for fourth down conversions, the distance to gain the first down and distance to the endzone, as well as the ratio of the two, are important in determining success.
- The models to predict WPA rely much more on the current game situation: the time remaining and point differential, etc.
- PlayCallR tends to make relatively aggressive play calls, which would increase game strategy and interest. In the fourth down situations included in the model (28,600 instances), the PlayCallR app recommends that the coach go for it 21.1% of the time. In real game play, coaches only went for the fourth down 15.4% of the time.
- Despite the score differential remaining relatively stable over the last 7 years, there has been an increase in aggressive play calling on fourth down in critical game situations.



Overall, the implementation of PlayCallR offers NFL teams an additional analytic tool that is objective and levels the playing field. The use of PlayCallR as one factor in the coach's decision-making process can lead to more aggressive play calling and increased competitiveness.

PAT PLAY CALLING RECOMMENDATIONS

The aggregated model will be pre-populated with the Offensive EPA/Play (Expected Points Added per Play: how effective is the offense on plays of under 5 yards), Defensive EPA/Play (Expected Offensive Points Added per Play: how effective is the defense on plays of under 5 yards), and the Kicker's Extra Point Make Percentage. The coach, in live game play, only needs to to input these two variables:

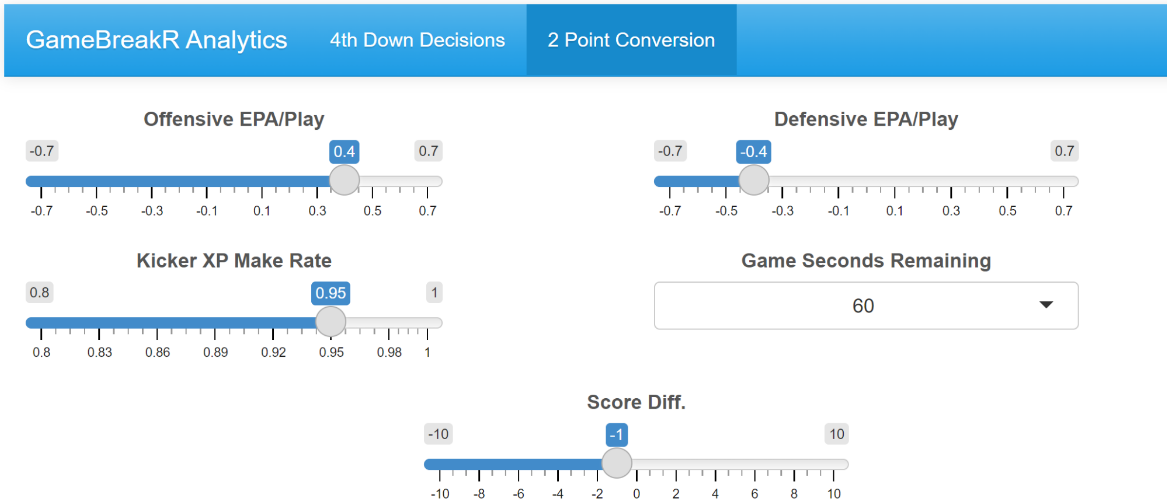
TIME REMAINING

POINT DIFFERENTIAL

The model then returns a recommendation to the coach as to whether they should go for a 2-point conversion or kick an extra point following a touchdown.

For example, our team is down by 1 point with only 1 minute remaining in the game. Our offense averages 0.4 points added per play within 5 yards, and the opponent's defense matches our competitiveness at -0.4 Defensive EPA/Play within 5 yards. Our kicker is fairly consistent, with a 95% accuracy rate. We just scored a touchdown, and the PAT model is recommending that the coach go for the two-point conversion instead of kicking the extra point.

playcallR



Go For 2

Down -1, 60 seconds left

Label	WPA if Success	WPA if Failure	Exp. Conversion Rate	Expected WPA
Go For 2	20.7	16.5	48.0	**18.5**
Kick XP	-1.0	-9.0	96.7	**-1.2**

WPA if Success: The estimated win probability added if the play is successful
WPA if Fail: The estimated win probability added if the play fails
Exp. Conversion Rate: The expected conversion rate for that decision
Expected WPA: The aggregate of success and failure - the projected WPA if that decision is chosen

Model Construction

The PAT play calling recommendation engine in PlayCallR is an aggregate of 6 different models using the variables listed in **Appendix C**:

- A logistic regression model to predict extra point conversion rate
- A logistic regression model to predict two-point conversion rate
- Two random forest models to predict WPA--both with successful and unsuccessful two-point conversion attempts
- Two random forest models to predict WPA--both with extra points made and missed

All 6 models are then used to calculate two figures, the “go for it” WPA and the extra point attempt WPA. In the example shown above, this is calculated as follows:

Go For 2

$$0.480\left(P_{success}\right) * 20.7(WPA) + 0.520\left(P_{failure}\right) *- 16.5(WPA) = 1.4\,WPA$$

Kick Extra Point

$$0.967\left(P_{success}\right) *- 1.0(WPA) + 0.033\left(P_{failure}\right) *- 9.0(WPA) =- 1.2\,WPA$$

The difference between the outputs is **<u>2.6 higher for "Go for 2"</u>**, which feeds the recommendation in PlayCallR.

Note that PlayCallR will return a recommendation of "toss up" if the WPA differential is greater than zero but under 1.8 percentage points, leaving this play call to the coach's discretion. The cutoffs for recommendations were based upon the quartile values of the WPA differential.

2 Point Conversion	Extra Point
52.6%	95.8%

R^2 Values

	2 Point Conversion	Extra Point
Success	0.9125	0.1415
Failure	0.8634	0.1311

The accuracies of the extra-point models are extremely high when compared to the two point conversion models. The model may not really add much value for extra points as the rate of success is 95.9%. So if a model predicted the extra point would always be made, the model would already be 95.9% accurate.

Two-point conversions are a bit more difficult to predict. This could be due to many factors, most notably that game flow and time remaining have a significant impact on two-point conversions. For example, if a team has been down all game, it is because they are not very good at generating points from within 5 yards from the endzone and the opposing team does not allow very much yardage in this tight window of space. Therefore, they would not be expected to do very well on a two-point conversion. However, based on our analysis, extra points are independent of all other variables and the kick just comes down to whether or not the kicker makes it. Two-point conversions have just a slightly better prediction rate than a coin flip.

Interestingly, the random forest models for two-point conversions do a very good job of accounting for the variability in WPA.

Insights

- The two most predictive factors in whether or not a team is successful on a two-point conversion are the offensive expected points added on plays within the 5 yard line and defensive expected points added on plays within the 5 yard line. It's fairly intuitive that prior success in short-yardage situations would be predictive of two-point conversion attempt success as two-point conversion plays begin on the 2 yard line.
- The only factor predictive of extra point success is the prior extra point success rate. We did isolate this by season as the length of the extra point kick increased from 20 to 33 yards in 2015.
- In the four random forest models that predict WPA (win percentage added), the game situation is much more critical, with the most important variables being time remaining and the point differential. This makes sense as most two-point conversion attempts take place late in close games.
- The aggregated model is fairly aggressive in routinely suggesting a two-point conversion attempt instead of an extra point kick. Because extra points are successful more than 95% of the time, it becomes a known option. The two-point conversion has around a 50/50 chance of success so both the risk and resulting reward is greater.

More advanced model types like neural networks including convolutional layers or recurrent layers for image processing were also considered. Published on June 6, 2021, Ben Baldwin demonstrated high success in image processing on NFL coverage classification using CNNs. Ben treats player tracking data as an image recognition problem and then applies well established computer vision techniques. The model achieved around 78% accuracy. Image processing however requires a large amount of calculation and the processing time simply cannot support the fast decision making nature of this project. The direction of the image recognition product, however, is not in-line with our project goal. We aim to provide the best decision given the situation at hand instead of predicting what a specific coach would do – that might be an impulsive decision and not the best one for winning, and that is not what we want.

In addition to the logistic regression models and the random forests, we also ran a boosted gradient model to evaluate vs. the random forest. In the end, the boosted gradient was not more accurate than the random forest and processing time was much slower. Given the critical nature of speed in NFL play calling, we chose to use the random forest. The random forest really offers the idea of "wisdom of the crowd" in that it's the accumulation of hundreds of independent model trials.

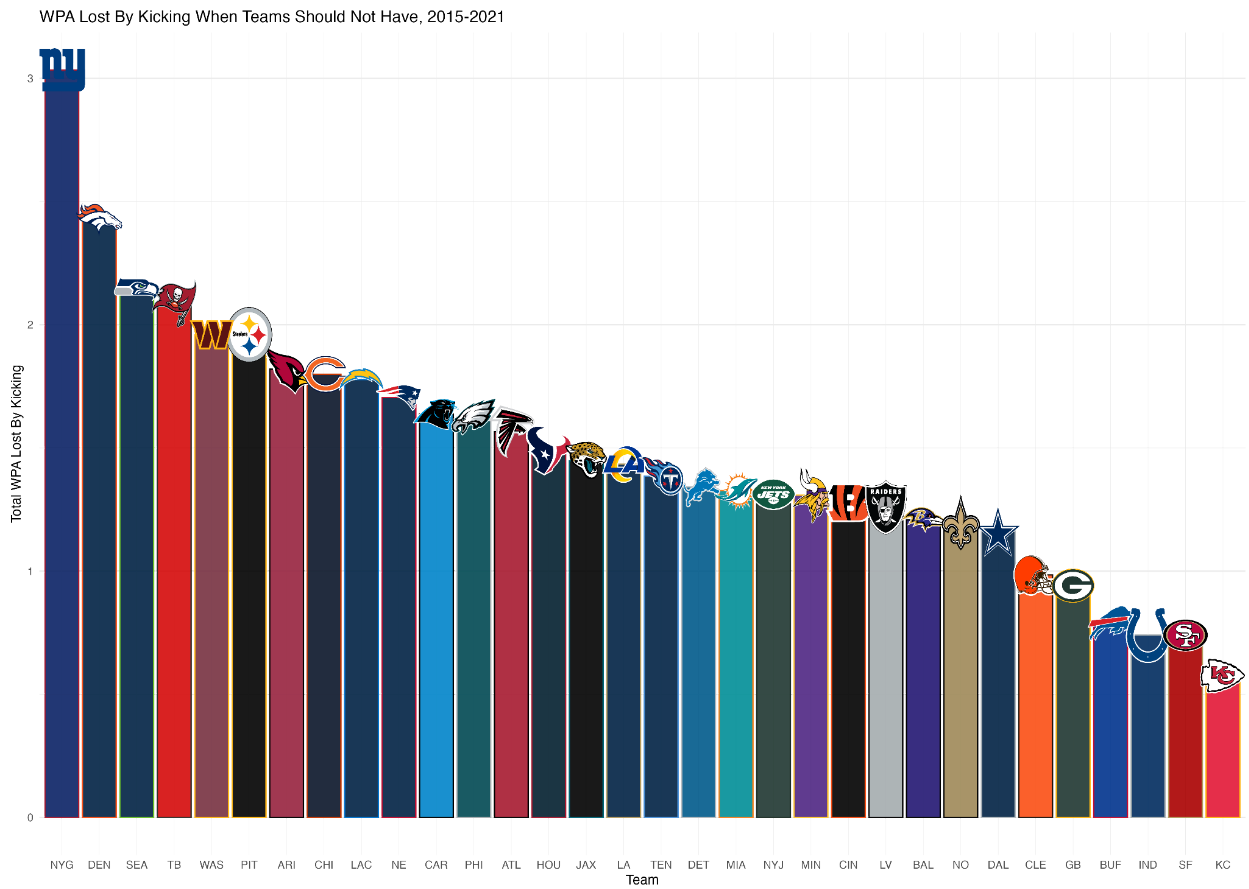
Using Principal Component Analysis to reduce dimensionality was also considered, yet we have not achieved much improvement. We did not observe highly correlated variables and thus using PCA resulted in significant color loss. Even though random forest is a blackbox model and utilizing such a model undoubtedly sacrifices the ability to explain the model rationale to some degree, we still hope to gain basic interactive insights from the variables used. Applying PCA to combine variables proved to be challenging to explain input variables to potential coaches and product users. Secondly, if the model is trained by PCA process variables, we will have to set an extra pre-processing stage to prepare PCA before raw data is pumped through our model, which can create difficulties for different client's data pipelines and increase the speed of real-time model decision output. Given that the improvement from using PCA was minimal, PCA was dropped out of our model input construction.

CONCLUSIONS

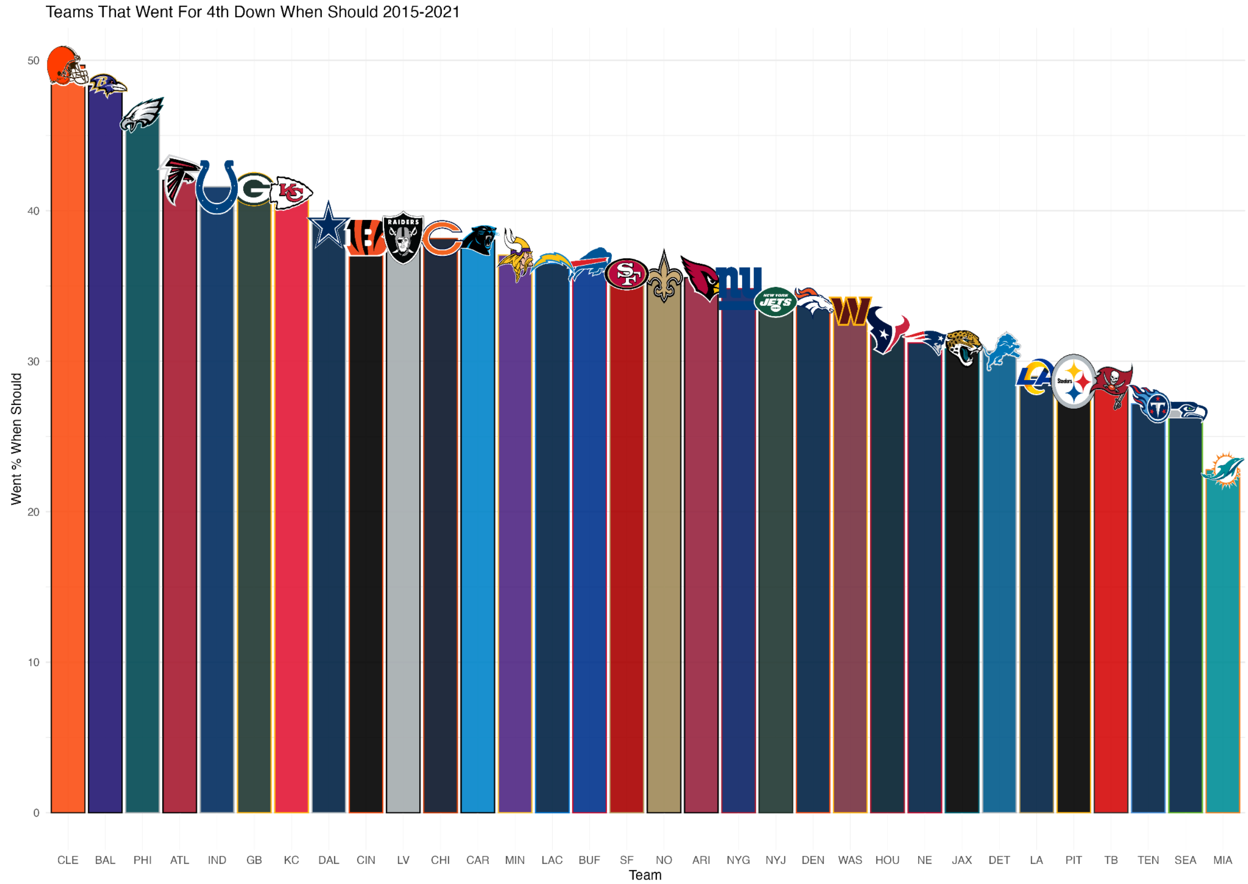
The fourth down model produced actionable recommendations that the coaches can utilize. The model makes more aggressive play call recommendations than the coaches generally make. Utilization of this model across all 32 NFL teams would result in more competitive games, generating increased fan interest and advertising revenue. It also offers the opportunity to level the playing field for teams without robust analytics teams.

If we view the last 7 years of fourth down play calls, we see that teams have very different trendlines.

The graph below shows that the New York Giants lost the most points and the Kansas City Chiefs lost the fewest points in WPA by kicking field goals when they should have gone for the 4th down conversion. So, the Giants were unsuccessful by being less aggressive than the model's recommendations. The Chiefs lost hardly any points in WPA because they much less frequently opted to kick a field goal when they should have attempted the fourth down conversion.



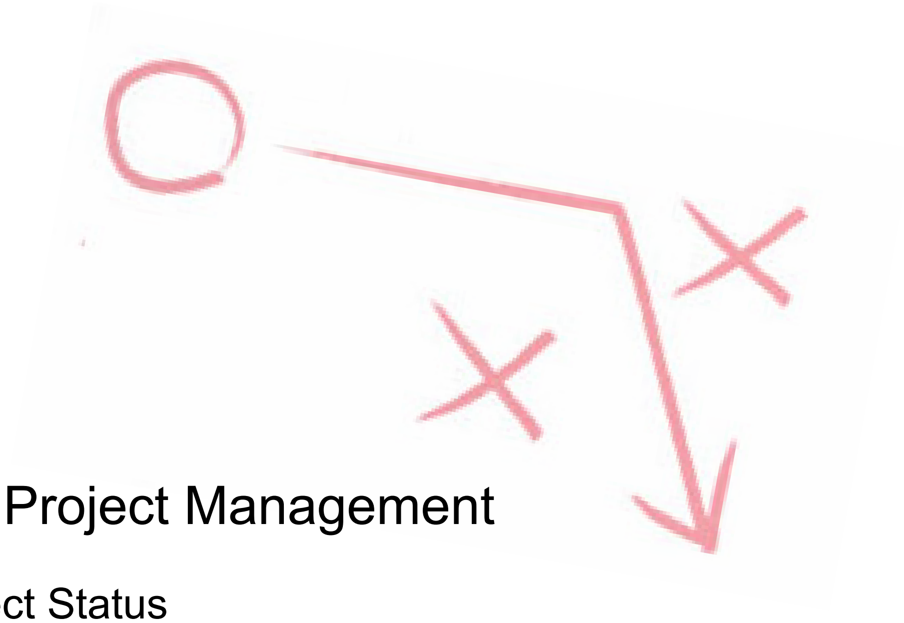
The second chart ranks which teams “went for it” on fourth down when the model recommended they should. So the Cleveland Browns’ fairly aggressive play calling matched the model’s recommendation most often of any team over the last 7 sevens. This is partially due to more recent aggressive play calling, but in the few years prior to that, the Browns were often down by large margins late in games. In those situations, teams almost always “go for it” on fourth down.



Both models, fourth down and PAT, utilize a calculated metric derived from 6 individual models to define a recommended play call. The fourth down model has more recommended outcome options (go for it, kick a field goal, or punt) and also more predictive variables than the PAT model. In the end, the fourth down model will likely be significantly more useful to the NFL teams than the PAT model.

The PAT model really only has two predictive variables, the score differential and the kicker effectiveness. Because the kickers so often make the extra point, the model almost always suggests the team make the more aggressive play call and “go for 2.” Because both the risk and reward is greater on the two-point conversion attempt, the model will recommend going for it.

We strongly believe that implementation of the fourth down model app PlayCallR for all 32 teams would greatly benefit the NFL, the individual teams, and the fans.

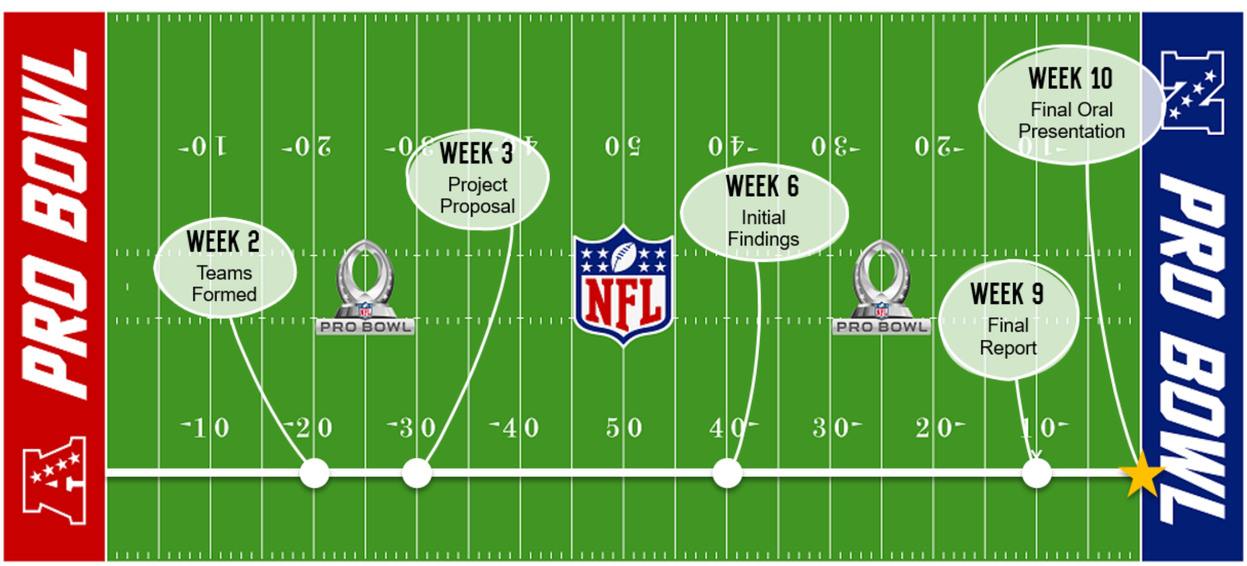


Q4: Project Management

Project Status

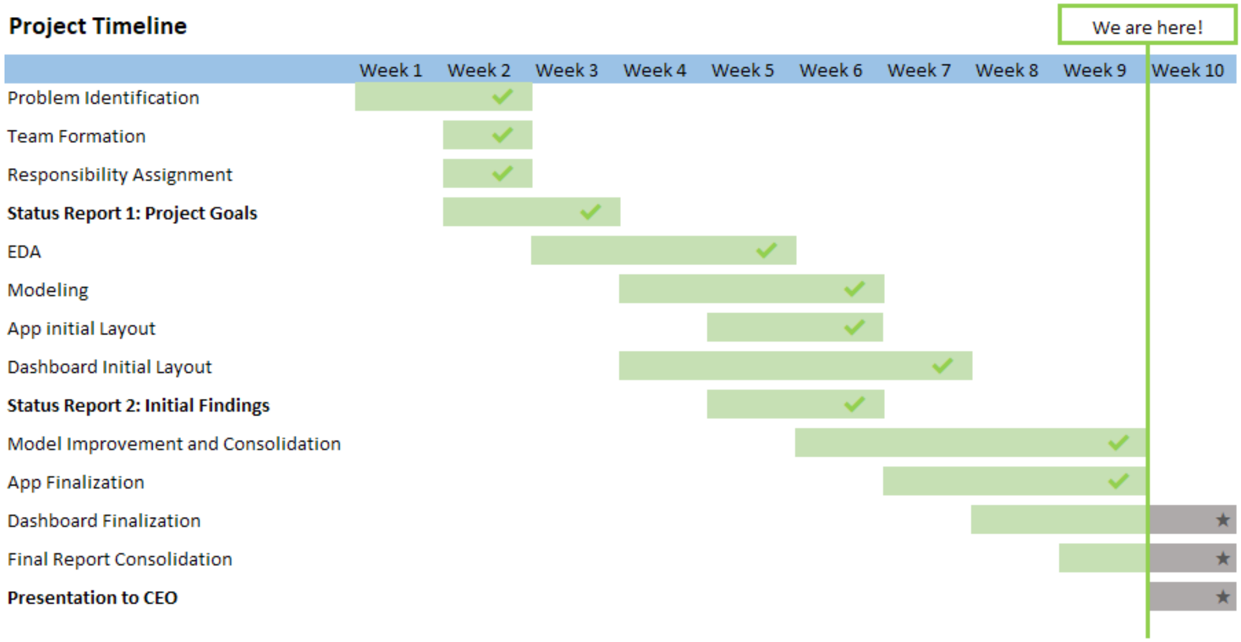
The GameBreakR Analytics team has completed all modeling for the fourth down conversions and PAT models, an R Shiny app, as well as an app within Tableau.

Overall Timeline



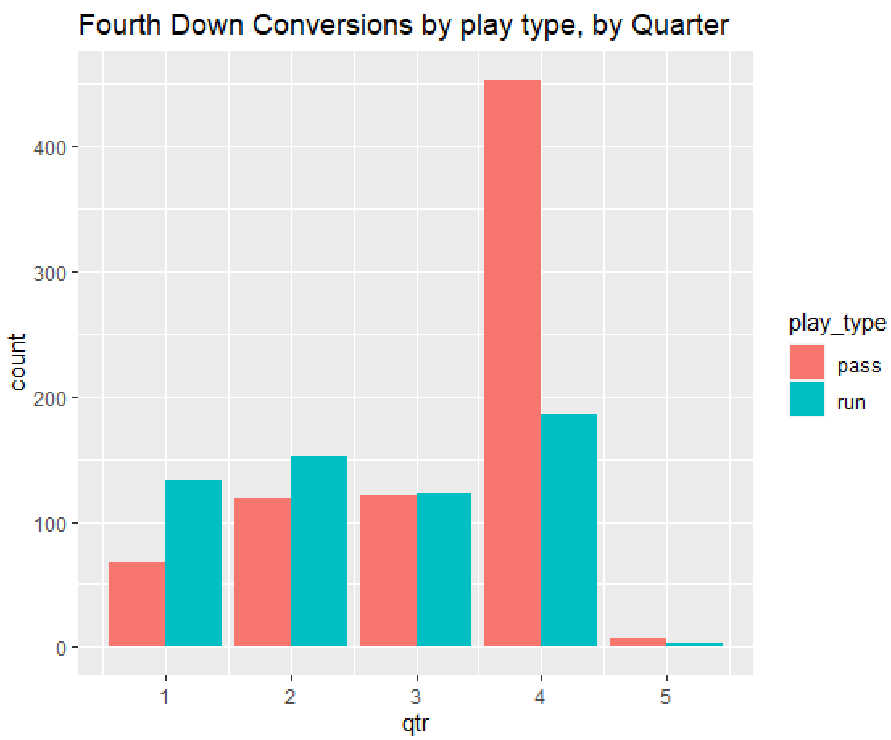
Detailed Gantt Chart

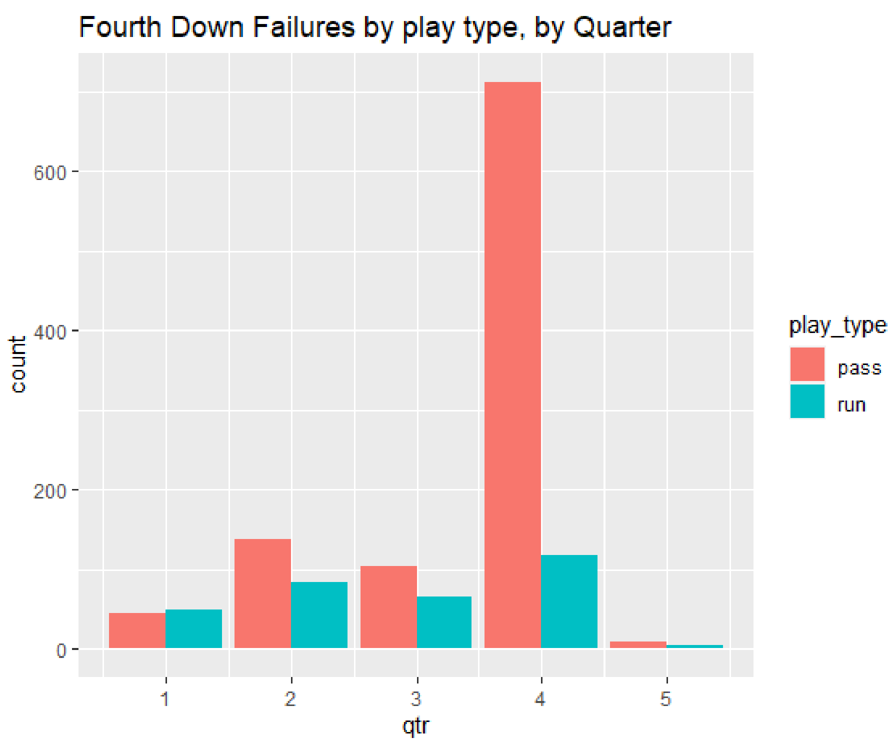
Project Timeline

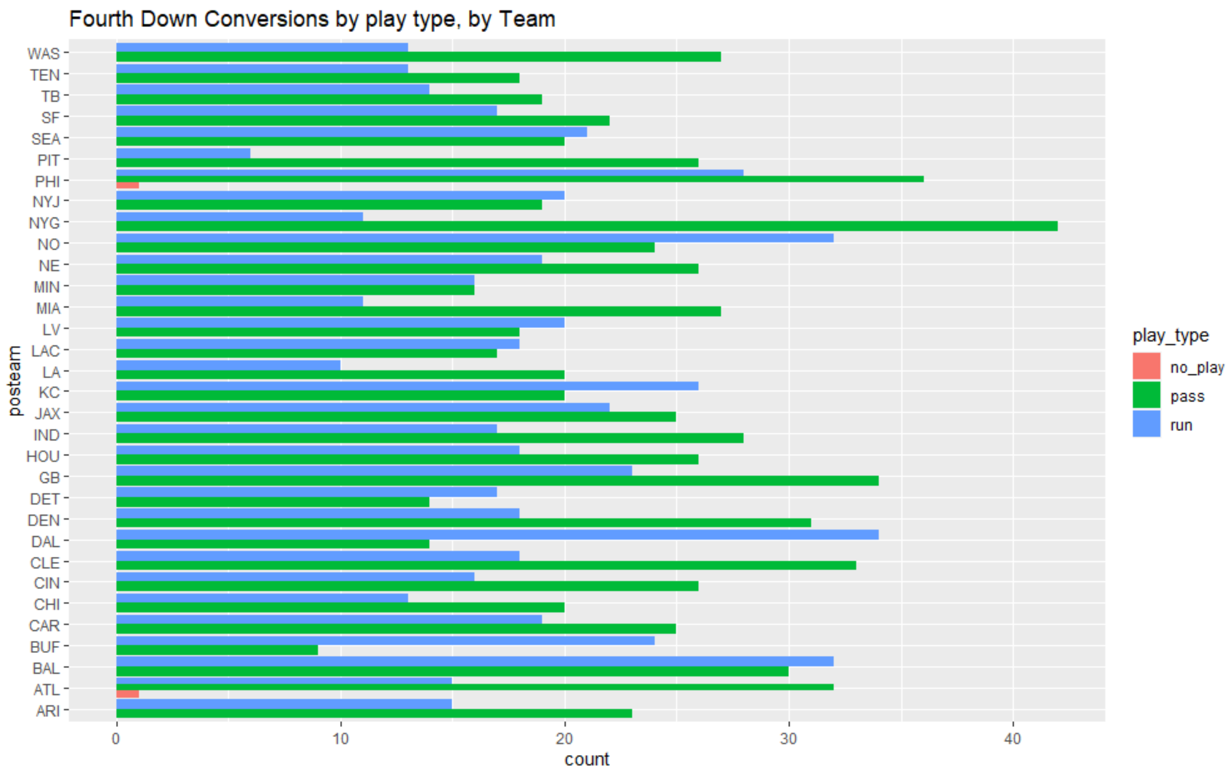


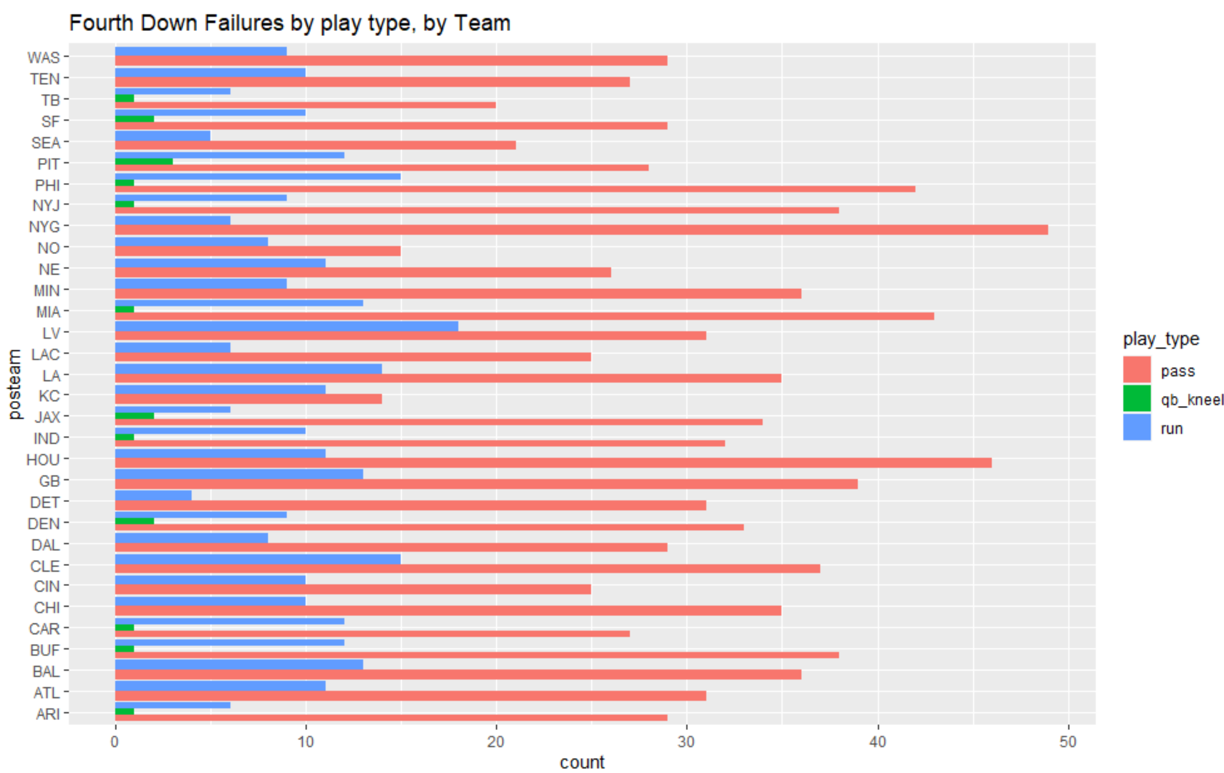
Overtime

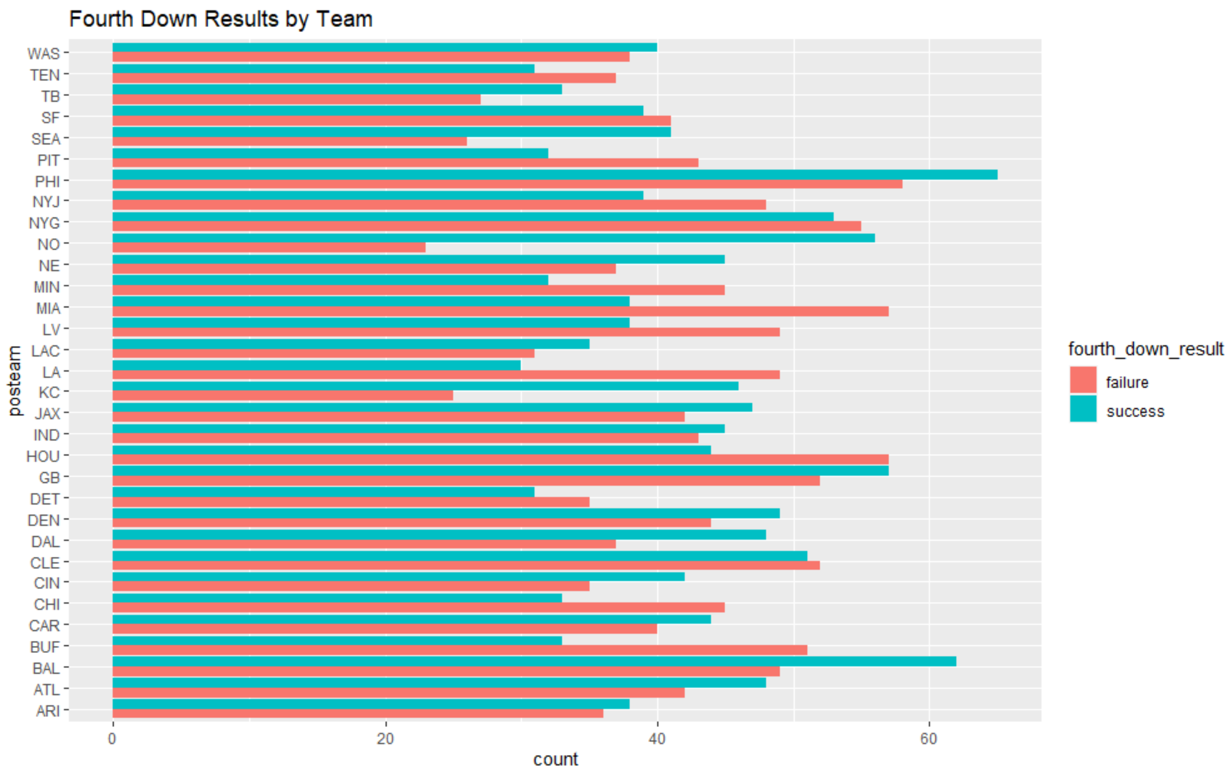
Appendix A: EDA Graphs for fourth down Data


Fourth Down Conversions by play type, by Quarter
400
300
200
100
0
count
1
2
3
4
5
qtr
play_type
pass
run

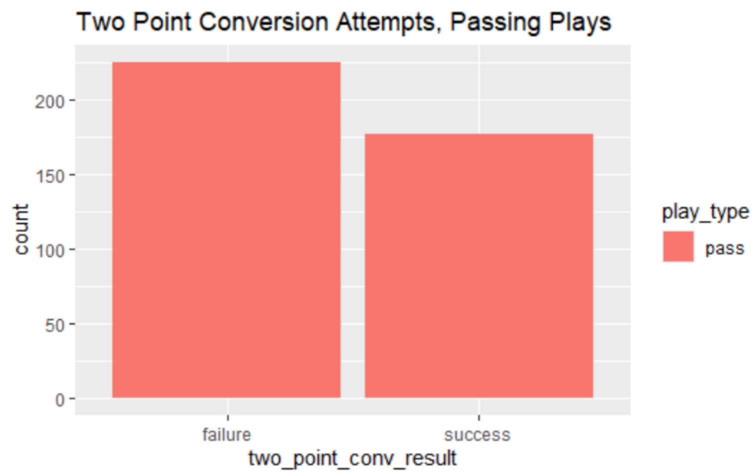

Fourth Down Failures by play type, by Quarter
600
400
200
0
count
1
2
3
4
5
qtr
play_type
pass
run

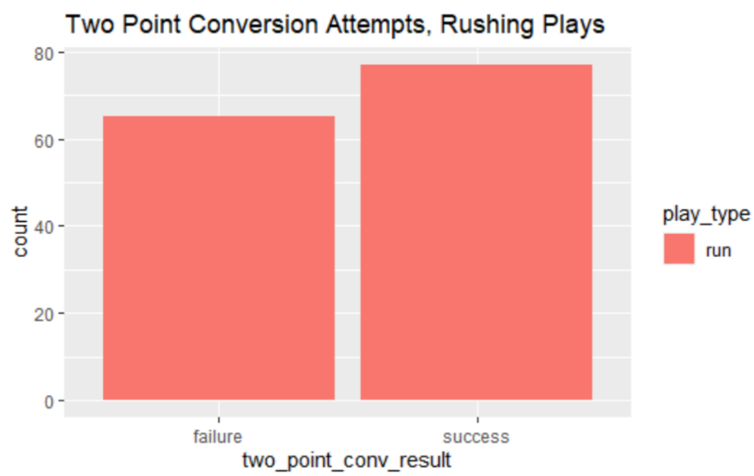

Fourth Down Conversions by play type, by Team
posteam
WAS
TEN
TB
SF
SEA
PIT
PHI
NYJ
NYG
NO
NE
MIN
MIA
LV
LAC
LA
KC
JAX
IND
HOU
GB
DET
DEN
DAL
CLE
CIN
CHI
CAR
BUF
BAL
ATL
ARI
0
10
20
30
40
count
play_type
no_play
pass
run

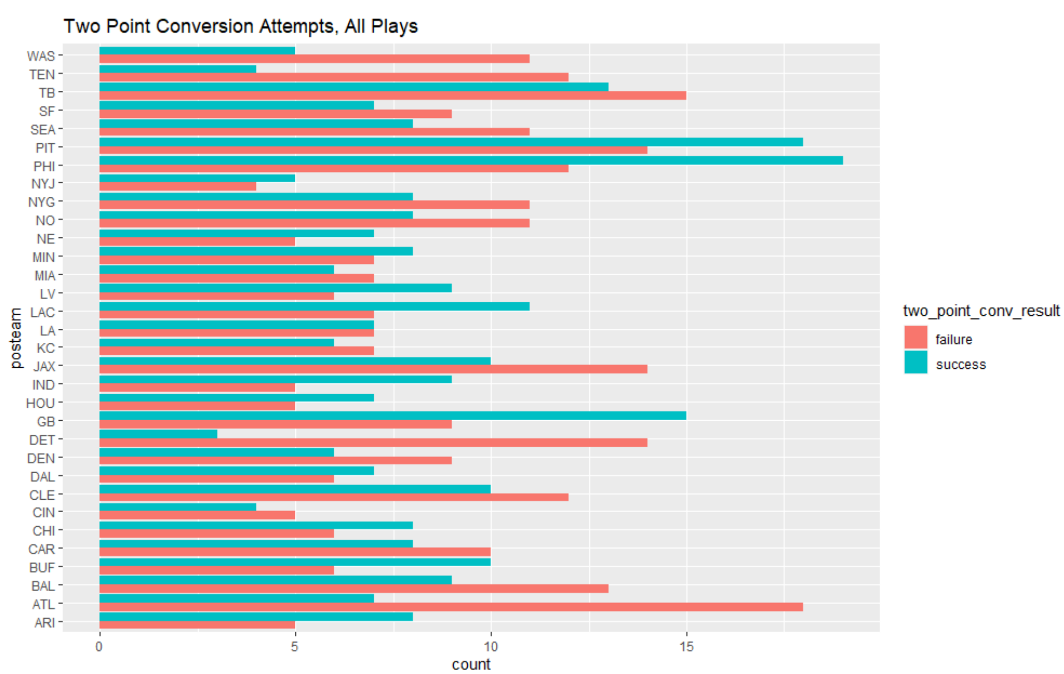

Fourth Down Failures by play type, by Team
posteam
WAS
TEN
TB
SF
SEA
PIT
PHI
NYJ
NYG
NO
NE
MIN
MIA
LV
LAC
LA
KC
JAX
IND
HOU
GB
DET
DEN
DAL
CLE
CIN
CHI
CAR
BUF
BAL
ATL
ARI
0
10
20
30
40
50
count
play_type
pass
qb_kneel
run


Fourth Down Results by Team
WAS
TEN
TB
SF
SEA
PIT
PHI
NYJ
NYG
NO
NE
MIN
MIA
LV
LAC
LA
KC
JAX
IND
HOU
GB
DET
DEN
DAL
CLE
CIN
CHI
CAR
BUF
BAL
ATL
ARI
posteam
0
20
40
60
count
fourth_down_result
failure
success

Appendix B: EDA Graphs for PAT Data


Two Point Conversion Attempts, Passing Plays
count
200
150
100
50
0
failure
success
two_point_conv_result
play_type
pass


Two Point Conversion Attempts, Rushing Plays
count
80
60
40
20
0
failure
success
two_point_conv_result
play_type
run


Two Point Conversion Attempts, All Plays
posteam
WAS
TEN
TB
SF
SEA
PIT
PHI
NYJ
NYG
NO
NE
MIN
MIA
LV
LAC
LA
KC
JAX
IND
HOU
GB
DET
DEN
DAL
CLE
CIN
CHI
CAR
BUF
BAL
ATL
ARI
0
5
10
15
count
two_point_conv_result
failure
success

Appendix C: Variable Definitions

Project Variables - Fourth Down Models

	Type	Description	4th Down Conv. Rate Regress	FG Conv. Rate Regress	4th Down Success RF	4th Down Fail RF	FG Made RF	FG Missed RF
INDEPENDENT VARIABLES (ranked by importance)								
yardstogo	integer	Yards to first down	2		4	4	4	4
yardline_100	integer	Yards to endzone	1	1	3	3	3	3
yardstogo:yardline100	float	Ratio	3					
game_seconds_remaining	integer	Seconds left in the game			2	2	1	1
score_differential	integer	Points ahead or behind			1	1	2	2
DEPENDENT VARIABLES								
converted	binary	Made first down 1/0	X					
fg_made	binary	Made field goal 1/0		X				
WPA	float	Win percentage added			X	X	X	X

Project Variables - Points After Touchdown Models

	Type	Description	Two Point Conv. Regress	Extra Point Conv. Regress	Two Point Success RF	Two Point Failure RF	Extra Point Made RF	Extra Point Missed RF
INDEPENDENT VARIABLES (ranked by importance)								
off_epa_in_5	float	Offensive expected points added with 5 yards of endzone	1		4	4		
def_epa_in_5	float	Defensive expected points added with 5 yards of endzone	2		3	3		
game_seconds_remaining	integer	Seconds left in the game	3	3	2	2	1	1
score_differential	integer	Points ahead or behind	4		1	1	2	2
xp_make_rate	float	Extra point success rate by season		1				
season	integer	Year (NFL season)		2				
DEPENDENT VARIABLES								
success	binary	Made two point conversion 1/0	X					
xp_made	binary	Made extra point goal 1/0		X				
WPA	float	Win percentage added			X	X	X	X

Appendix D: About Us

GameBreakR Analytics aims to bring objective statistical analysis to all teams in the National Football League. Our goal at **GameBreakR** is to provide NFL teams with insights that analyze decision-making to help inform their gametime strategies.

Meet the Team

Kristen Allen, **Project Manager**
Kristen has 20+ years of experience in the marketing field, primarily in digital marketing/eCommerce. She currently works for an outdoor recreation products manufacturer.

Shruthi Harve Iyengar, **Data Visualization Expert**
Shruthi has spent over 6 years in business intelligence and finance reporting roles in the media industry. She also dabbles in web development and creative fields on the side. She currently analyzes content performance for Peacock.

Max Pauly, **Modeler/Analyst**
Max has 4 years of analytics experience in the retail industry, most recently for Gap, Inc. He currently works as a Senior Data Analyst on the Advanced Analytics team, supporting inventory management process improvement.

Michael Venit, **NFL Domain Expert**
Michael spent 4 years working for the Los Angeles Rams in scouting operations and talent evaluation. He's also spent two years providing analytics on a volunteer basis to two D1 NCAA football teams.

Alex Zhou, **Researcher**
Alex has 4 years of analytics/data science experience in the storage and logistics industry. He was also a TA in the Data Science program at the University of Denver.

Contribution by Member

	Domain Expert	Data Visualization	Project Management	Research	Modeling
Kristen Allen		○	●	○	
Shruthi H. Iyengar	○	●	○		○
Max Pauly	○			○	●
Michael Venit	●	○	○		○
Alex Zhou		○		●	○