Final Report: First Down or Score Classification Analysis

Introduction:

The National Football League (NFL) is an organization that has embraced technology and data science to improve its games, make informed decisions during play calling, and enhance the overall experience for its fans and players. The NFL's Big Data Bowl is an annual competition that challenges data scientists to use advanced analytical methods to extract insights from the vast amount of data generated by the league.

In this paper, I present a data science project focused on measuring the chance of converting a play into a first down or scoring a touchdown in the NFL using the NFL Big Data Bowl 2022 dataset. I utilized machine learning algorithms to analyze the data and identify the most critical factors that impact a team's success in achieving these outcomes.

The NFL Big Data Bowl 2022 dataset provides detailed information about every play in every game of the first 8 weeks of the 2021 NFL season, including player tracking data, game context, and play outcomes. I preprocessed the data, created metrics, and extracted features relevant to player and team performance, game situations, and play types. We then applied machine learning techniques to build models that predict the likelihood of first down conversions and touchdowns based on these features.

This study contributes to the growing body of research on data science and sports analytics, demonstrating the value of using advanced analytical methods to extract insights from complex and diverse data sources. The findings have implications for coaches and players who can use these insights to optimize their playcalling and strategy. Additionally, our project provides a foundation for future research and development in the field of sports analytics.

Data:

The NFL Big Data Bowl 2022 dataset includes play-by-play data, which is further broken down into frame data for each play. The frame data captures the movement and positioning of players on the field at 10 frames per second, providing a granular level of detail about each play. To analyze the data, I processed it into play-by-play data while retaining important features related to play performance, game context, and play outcomes.

The preprocessing pipeline included several steps, such as aggregating the frame data into summary statistics for each play, and encoding categorical variables (e.g., play type, offensive formation) using one-hot encoding. I also calculated several new features based on the available data, such as the distance to the first down marker or touchdown line (whichever is closer) and routes run by a player during the play.

After preprocessing, I selected a subset of features that I believe are most relevant to predicting play outcomes. These features included player and team performance metrics (e.g., yards gained, receiver position relevant to closest defender), game context variables (e.g., down and distance), and play type indicators (e.g., type of route run by the receiver). I then used various machine learning algorithms, including logistic regression and random forests, to build models that predict the probability of a team converting a first down or scoring a touchdown.

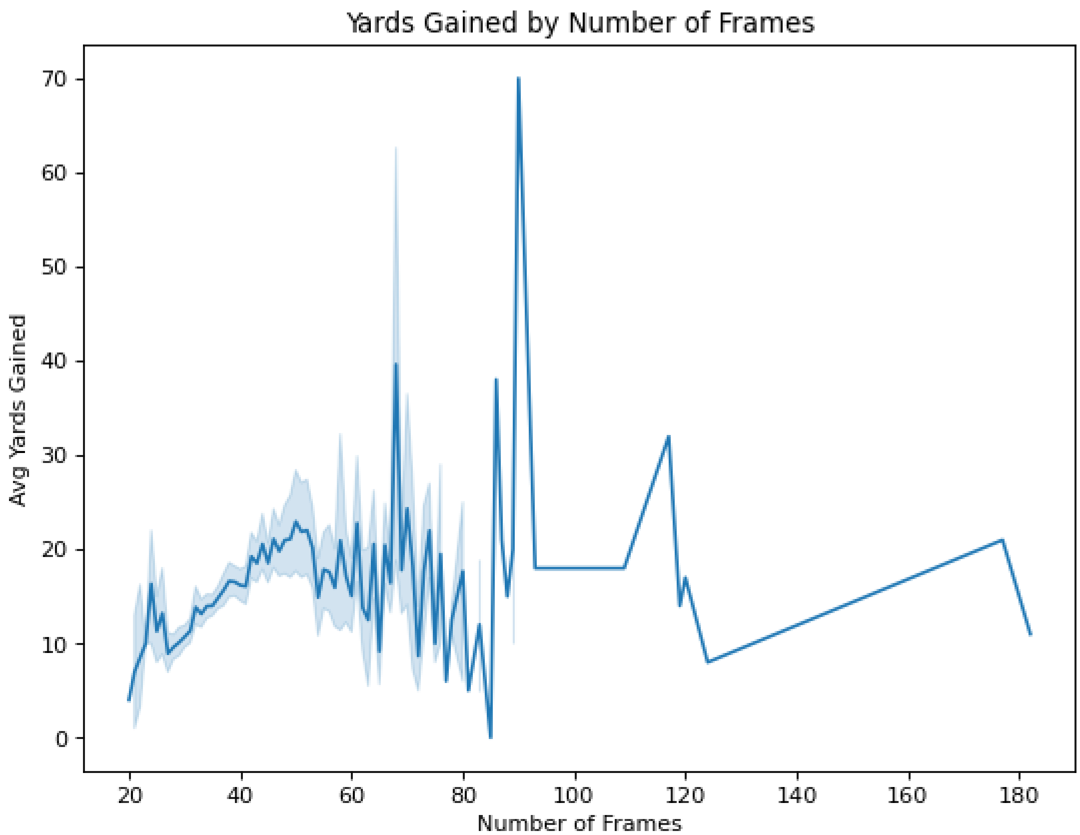

Fig 1. Yards Gained Broken down by Number of Frames

Metric Creation:

In this paper, I present some metrics that we created using this dataset. These metrics are designed to provide deeper insights into the game by analyzing key factors such as the separation between a receiver and a defender, the distance of a receiver from the ball, the type of receiving route used, and whether a first down was achieved.

To create these new metrics, I processed the raw frame data provided in the dataset. I then extracted and analyzed relevant features such as the location of the ball, the position of the players on the field, and the timing of the plays.

Metrics:

1. Separation from defender: This metric measures the distance between a receiver and the defender who is covering them. The distance is calculated at the moment the ball is snapped to the QB. This metric can be used to evaluate how successful a receiver is at creating space between themselves and their defender during their route. This, in turn, can lead to more successful pass plays, as it provides the quarterback with a greater target to throw to. A larger separation can indicate the offense is in need of a significant chunk of yardage to achieve a first down or score or that the receiver is too fast to be lined up close to the line of scrimmage.
2. Distance from the ball to the receiver: This metric measures the distance between a receiver and the ball when it is snapped. This metric combined with the other metrics can evaluate whether the route run by the receiver is best for someone lined up in that position and whether positioning closer to or further from the ball impacts the probability of yardage gained by the receiver. If the distance is small, it indicates that the receiver is closer to the ball (e.g. receiver is a tight end, goal-line play). If the distance is large, it indicates that the receiver is further positioned from the ball (e.g. player is on the outside). This metric can be used to identify potential issues with timing, accuracy, or positioning in the passing game.
3. Catching receiving route: This metric categorizes the type of receiving route used by a receiver on a given play. This can include routes such as slants, posts, or outs and whether the route was short (0-15 yds), medium (16-30 yds), or long (30+ yds) This metric can be used to evaluate the effectiveness of different routes and to identify patterns in the performance of getting a first down or a score. For example, if a receiver route consistently performs well on a certain type of play, the team may choose to incorporate that route more frequently in their game plan.
4. First down or Score achieved: This metric indicates whether a first down was achieved on a given play. This metric can be used to evaluate the success of a team in moving the ball down the field and converting on critical downs. If a team consistently achieves first downs, it indicates that they are successful at sustaining drives and keeping possession of the ball. If a team struggles to achieve first downs, it may indicate weaknesses in their offensive strategy or execution. This metric can be used to evaluate the effectiveness of different offensive strategies and to identify patterns in-game performance.

Exploratory Analysis:

This section of the paper focuses on examining the NFL Big Data Bowl 2022 dataset and exploring the relationships between the features and the target variable - whether a first down or touchdown will occur. The goal of this analysis is to gain insights into the data and identify potential patterns and relationships that can be used to develop a new predictive metric for first down or touchdown occurrences. By exploring the data, we can identify important features and gain a deeper understanding of the factors that contribute to a successful offensive play in the NFL. This EDA section will provide a detailed examination of the data, including the distribution of the features, the relationships between the features and the target variable, and any potential outliers or anomalies in the data. This analysis will serve as a foundation for the development of a new predictive metric that can be used to evaluate offensive performance in the NFL.

First, I will examine the impact of offensive formation on the result of offensive plays in the NFL. Findings suggest that offensive formation does not have a large impact on play result over the course of a season[Fig 2]. Specifically, we found that the Empty Formation had a mean of 6.56 yards gained on play and a standard deviation of 9.86. The Shotgun Formation, which was the most common formation in the dataset, had a mean of 6.26 yards gained on play and a standard deviation of 9.9. The Singleback Formation had a mean of 7.66 yards gained on a play, but a larger standard deviation of 11.49.

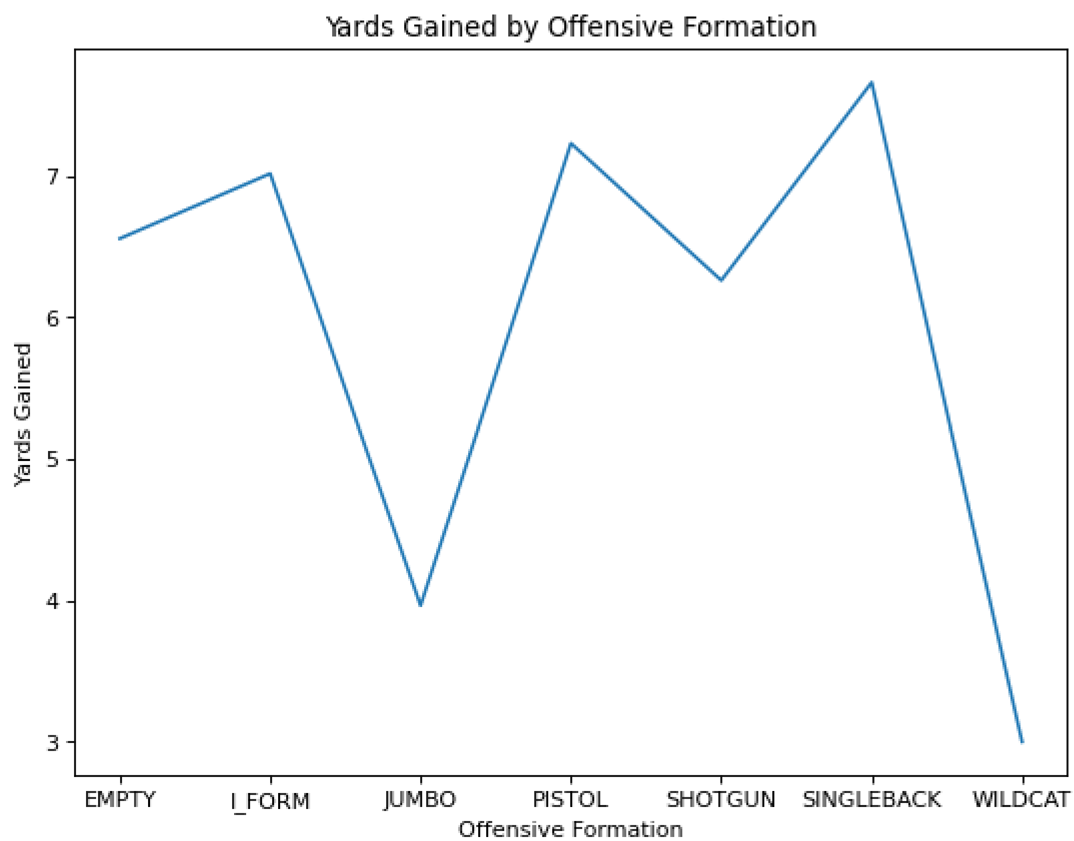

Fig 2. Average Yards Gained on a play by Offensive Formation

Second, I decided to look into how defensive pass coverage affects play results. The analysis suggests that pass coverage played by the defense does not have a large impact on play results over the course of a season. The most common play types used in order are Cover-3, Cover-1, and Cover-2 with Quarters a close fourth. Among these common coverages, the difference in mean between the largest mean yards gained and the smallest mean yards gained is about 0.2 yards. This indicates that pass coverage does not have a significant impact on yards gained on play. In addition, I found that the standard deviations for yards gained on play for each coverage type were relatively large (ranging from 9.5 yards to 11.5 yards), indicating significant variation in play results within each coverage type. The 2-Man coverage gives up the most yards on average (only 194 plays) but also has the largest standard deviation at 12.45 yards. This suggests that other factors, such as the specific play called or the skill of the individual players involved, may have a larger impact on play results than pass coverage alone. It also makes sense that the red zone plays have the lowest average yards gained because the offensive team is close to the goal line and the range of available yards to gain is low.

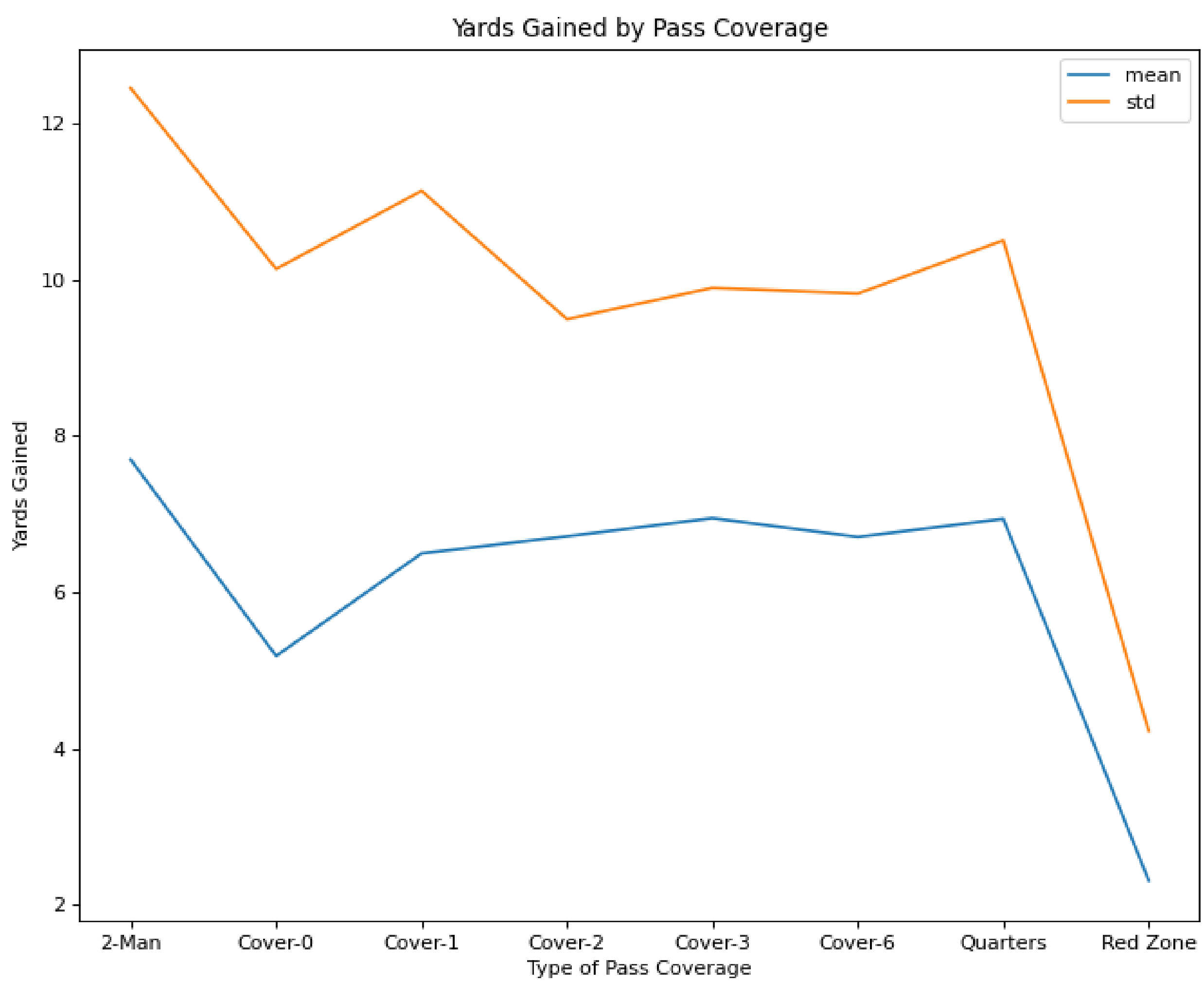

Fig 3. Average Yards Gained based on Defensive Pass Coverage

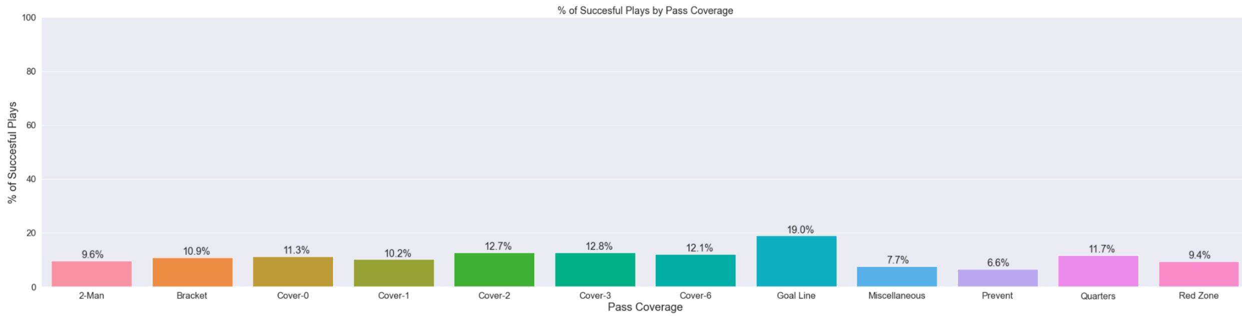

Fig 4. Percent of plays that lead to First Down/Score by Pass Coverage Type

I then decided to look into the breakdown of yardage gained by both defensive pass coverage and offensive formation. The most interesting tidbit is the disparity in average yards gained in Cover-2 defensive formation. With an ‘Empty’ (no RBs lined up in the backfield) offensive formation, the mean yards gained is right under 6 yards per play but for ‘Singleback’ (one RB lined up in the backfield with QB) formation, the mean yards gained is 8.6 - a difference of 43%. Of course, the standard deviation for the ‘Empty’ formation is significantly lower than the standard deviation for the ‘Singleback’ formation.

pff_passCoverage	offenseFormation	count	mean	std	min	25%	50%	75%	max
2-Man	SHOTGUN	164.0	7.371951	12.626856	-11.0	0.0	4.0	13.0	75.0
Cover-0	SHOTGUN	167.0	5.760479	10.597345	-11.0	0.0	0.0	7.0	50.0
Cover-1	EMPTY	321.0	6.174455	10.075377	-16.0	0.0	3.0	11.0	56.0
	SHOTGUN	1321.0	6.470855	11.349046	-18.0	0.0	1.0	11.0	91.0
	SINGLEBACK	232.0	6.637931	11.030376	-12.0	0.0	0.0	12.0	57.0
Cover-2	EMPTY	157.0	5.987261	8.942112	-10.0	0.0	4.0	9.0	61.0
	SHOTGUN	693.0	6.470418	9.123165	-34.0	0.0	5.0	10.0	73.0
	SINGLEBACK	141.0	8.609929	11.568290	-11.0	0.0	6.0	13.0	52.0
Cover-3	EMPTY	375.0	6.928000	9.148077	-13.0	0.0	6.0	10.5	57.0
	I_FORM	124.0	7.419355	9.911886	-10.0	0.0	4.5	12.5	38.0
	SHOTGUN	1511.0	6.444077	9.339549	-17.0	0.0	5.0	10.0	77.0
	SINGLEBACK	527.0	8.009488	11.312276	-24.0	0.0	6.0	13.0	84.0
Cover-6	EMPTY	119.0	7.310924	11.269238	-10.0	0.0	5.0	10.0	68.0
	SHOTGUN	553.0	6.462929	9.358741	-15.0	0.0	4.0	10.0	70.0
Quarters	EMPTY	221.0	8.058824	10.984335	-15.0	0.0	6.0	12.0	75.0
	SHOTGUN	640.0	5.946875	9.200445	-13.0	0.0	4.0	10.0	72.0
Red Zone	SHOTGUN	213.0	2.643192	4.192808	-12.0	0.0	0.0	6.0	13.0

Fig 5. Yards gained broken down by Pass Coverage and Offensive Formation

Next, I wanted to look at the impact the number of defenders in the box has on the number of yards gained on average. Interesting to note that more than 5 defenders in the box leads to higher average yards gained but also leads to higher standard deviation. But the most interesting chart is Fig 7, which shows the likelihood of gaining a first down or score increases as there are more defenders in the box. This could be explained due to more defenders in the box would mean fewer defenders in coverage of the receivers.

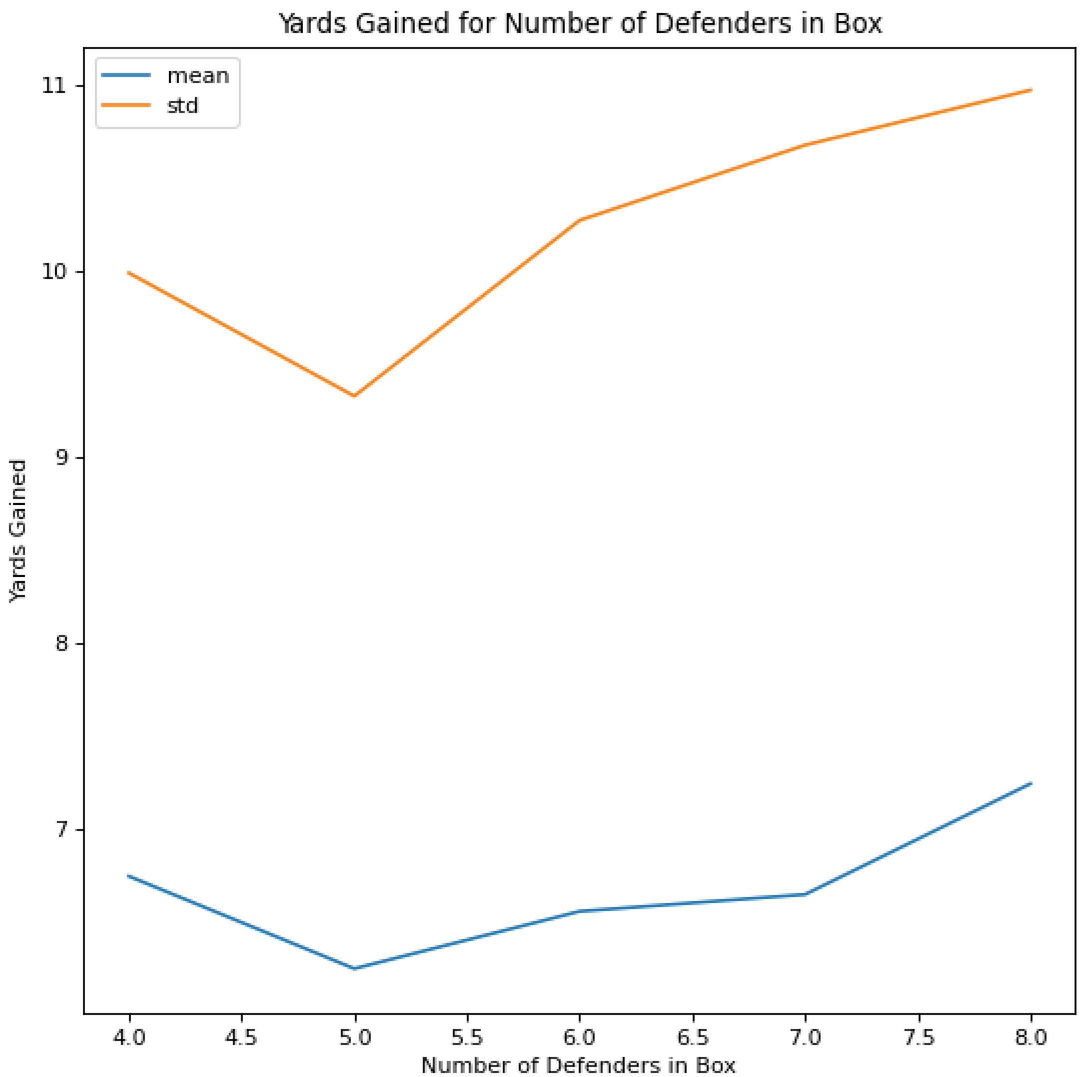

Fig 6. Yards gained broken down by the Number of Defenders in the Box

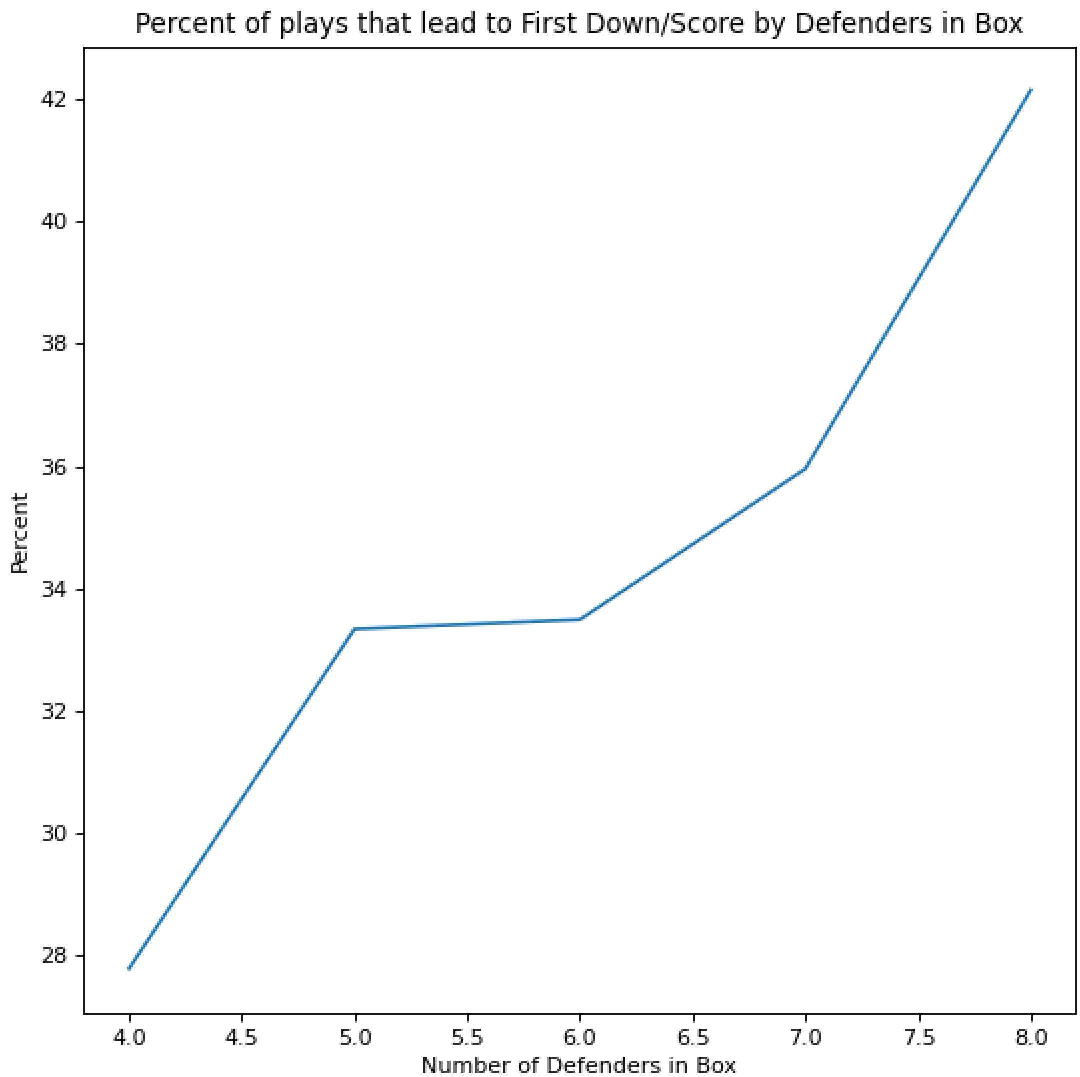

Fig 7. Percent of plays that lead to First Down/Score by Defenders in Box

In order to gain a better understanding of the relationship between pass coverage and the success rate of plays, I looked at the frequency of each type of coverage used by defenses and their corresponding success rates in converting plays into first downs or touchdowns.

Excluding Prevent coverage, which is typically used in end-of-game situations to prevent long passes, I found that most coverages have similar success rates at scale[Fig 8]. Among the most commonly used coverages, Cover-6 had the highest success rate at 13.7%, followed closely by Cover-3 at 12.8%. On the other hand, Prevent and 2-Man were the least effective at converting plays, both with a success rate below 10%.

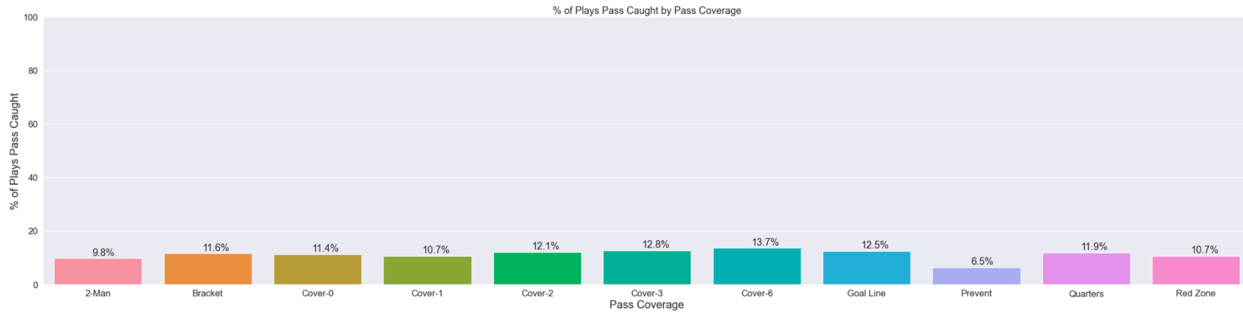

Fig 8. Percent of plays that lead to First Down/Score by Pass Coverage

Offensive Formation had a significant impact on the success rates of plays, as shown by Fig 9. The I-Form and Jumbo formation (albeit at a lower scale) had a success rate of over 17%, while the Pistol formation (easily the worst of the bunch) had a success rate below 10% at 8.9%. This finding suggests that offensive formation plays an important role in determining the outcome of a play, with certain formations being more successful than others. It is possible that some formations may provide better opportunities for the offense to make plays, while others may be more predictable and easier for the defense to defend against. This analysis shows the importance of offensive linemen for play outcomes as well. The I-Form and Jumbo formation are plays where the offensive line is heavily huddled to the QB and usually used in redzone or short yrdage situations. It is important to note that this analysis only provides a high-level overview of the relationship between offensive formation and play success. Further analysis is necessary to understand the underlying factors driving these trends and to determine whether other variables are also important predictors of play success.

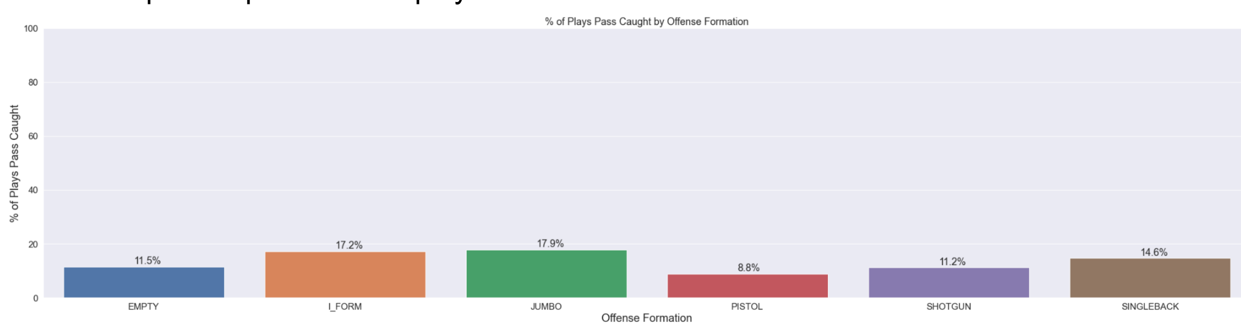

Fig 9. Percent of plays that lead to First Down/Score by Offensive Formation

Then, I wanted to dig into the breakdown of the position lined up for players that caught the ball[Fig. 10]. The analysis revealed that players lined up in the slot (left or right) caught the most number of passes, whereas WRs, TEs, and FBs lined up outside caught the least amount of passes. This finding could be attributed to several reasons, such as the ability of the QB to throw between the numbers more easily, or the fact that players lined up outside might be marked by better defenders. However, I was surprised to see such a high disparity between the number of passes caught by players in different positions.

After further research, the observation that slot receivers caught the most number of passes is consistent with the current trends in the NFL, where teams are increasingly using spread offenses and rely on shorter, quicker passes. Additionally, slot receivers typically have favorable matchups against linebackers or slot cornerbacks, which could also contribute to their higher catch rate. On the other hand, the lower number of passes caught by receivers lining up outside could be due to various reasons, including tougher matchups against cornerbacks, more difficult routes, or simply fewer opportunities due to game situations and play calling.

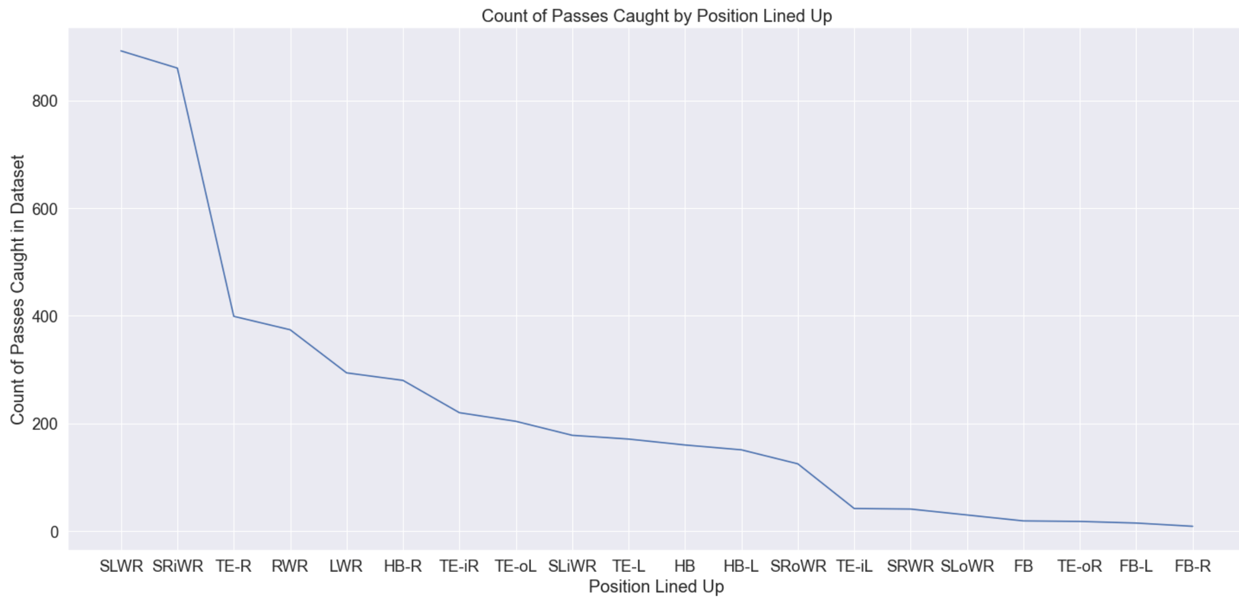

Fig 10. Count of Passes caught by players by Position Lined Up

As I was exploring trends within plays where the player caught the ball and found that the type of QB movement during the play had a significant impact on the success rates of plays. Specifically, when the QB had a designed rollout to the left or right side, the percentage of plays that led to a first down or score was notably higher than traditional dropbacks or scrambles. The success rates for designed rollouts were 18.4% for the left and 20.7% for the right[Fig. 11], compared to 12.8% for traditional dropbacks and below 10% for scrambles. Suggesting that QB movement during a play can have a meaningful impact on the success of the play. Designed rollouts may provide an advantage by changing the angle of the throw and creating more space for the receivers to work with. On the other hand, scrambles may be more difficult to execute successfully due to the unpredictable nature of the play.

Another possible explanation could be that the rollouts allow the QB to have a better view of the field and create more space for the receiver to get open. The lower success rate for scrambles could be due to the fact that it's a more improvised play and requires the QB to make split-second decisions. It would be worth further investigation to see if these trends hold up across multiple seasons, teams, and QBs, and whether they can be used to inform play-calling and game strategy.

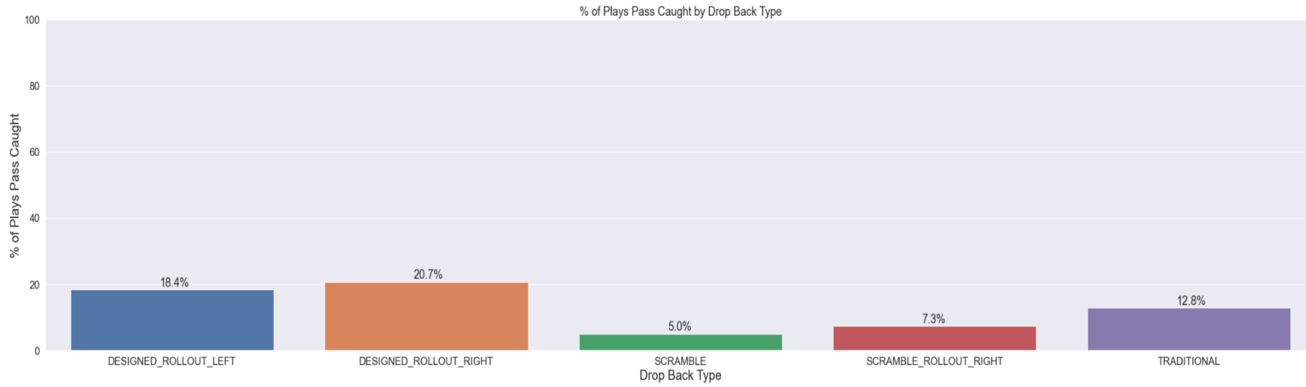

Fig 11. Percent of plays that lead to First Down/Score by QB Drop Back Type

Another metric I look at was the number of catch-eligible receivers and the success rate of the play[Fig. 12]. It's definitely possible that the number of catch-eligible receivers on a play is related to the success rate of the play. It could be that having more receivers increases the options for the quarterback and makes it harder for the defense to defend against, but at the same time, it may also require more precise timing and execution. Conversely, having fewer receivers may make it easier for the defense to anticipate the play, but may also allow for quicker, more efficient execution.

Also, it could be possible that the decrease in success rate with an increase in catch-eligible receivers could be due to the additional yardage required to gain a first down or score. It could also be related to the fact that with more receivers eligible to catch the ball, there are more defenders to cover them, making it more difficult to complete a pass. Another factor could be the complexity of the play design or the ability of the QB to read the defense and make the right decision with multiple options for receivers.

It would be interesting to further explore this relationship and possibly control for other factors such as defensive formations or offensive strategies.

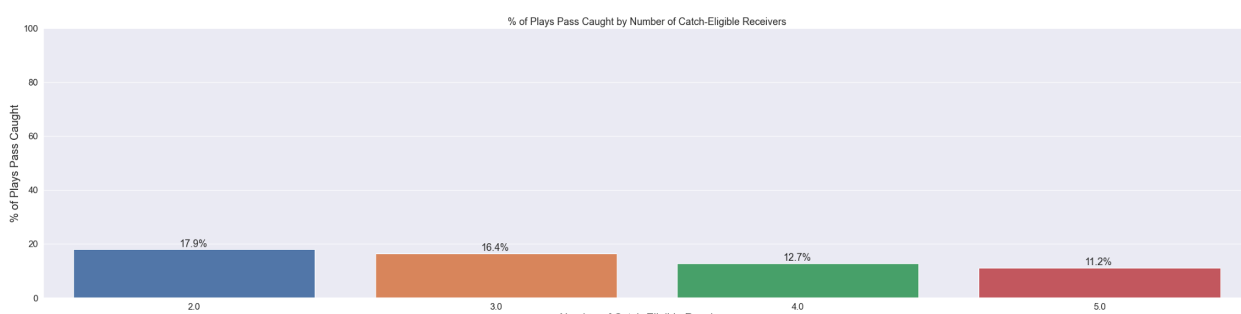

Fig 12. Percent of plays that lead to First Down/Score by Number of Catch-Eligible Receiver

Fig 13. below shows the breakdown of successful pass attempts by down with the 0th down counting as 2-PT conversion attempts. This is a very interesting finding! It's not necessarily intuitive that plays run on 4th down would have a higher success rate than those run on 3rd down, but it's possible that teams are more willing to take risks on when running plays on 4th down. It's also intriguing that plays run on 2nd down have the highest success rate, which may indicate that teams are able to catch the defense off guard more often on that down. This information could be useful for teams when deciding on play-calling strategies in different situations. It is worth further exploration as to why this is the case, as it could provide insights into offensive strategies or defensive weaknesses during specific downs.

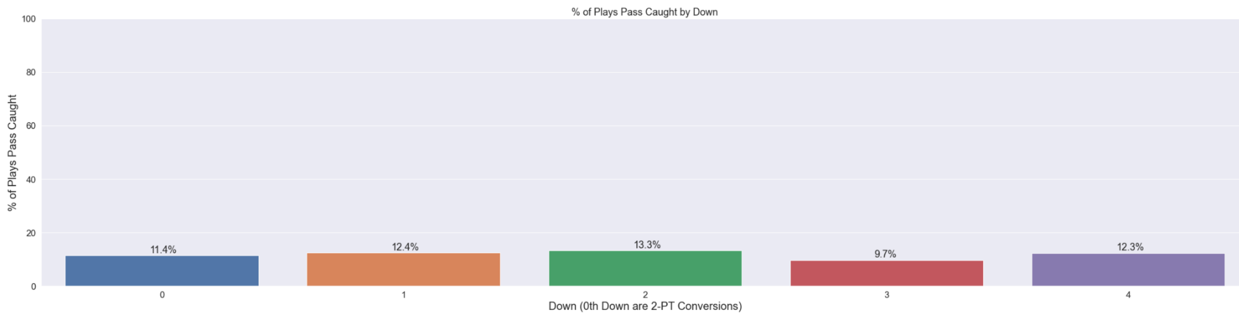

Fig 13. Percent of plays that lead to First Down/Score by Number of Catch-Eligible Receivers

In Fig 14 below, I break down the success rate by the quarter of play. It's interesting to see the variation in success rates across different quarters of the game. The higher success rate in overtime could be due to the higher stakes and pressure, leading to more focused and efficient play calling. It's possible that the success rates are influenced by factors such as game situation, fatigue levels, and defensive adjustments made by the opposing team throughout the course of the game, which is what led to the lower success rates in the 2nd and 4th quarters. It would be worthwhile to further explore these trends and potential factors that contribute to success rates across different quarters of the game.

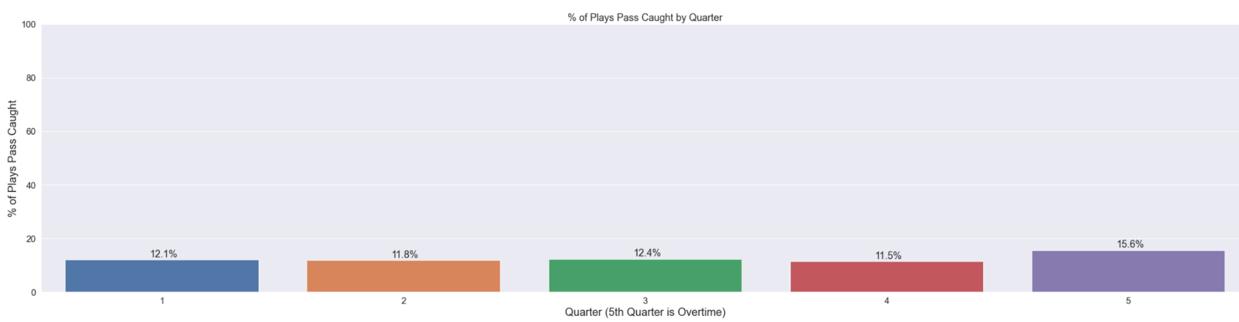

Fig 14. Percent of plays that lead to First Down/Score by Number of Catch-Eligible Receiver

Additionally, pass coverage type appears to have a slight impact on the success rate of achieving a first down or touchdown[Fig.15]. The success rate for achieving a first down or touchdown against man coverage is 10.7%, while the success rate against zone coverage is slightly higher at 12.6%. Interestingly, the success rate against "other" coverage types is slightly lower at 10.4%. However, it's important to note that this analysis only takes into account the impact of pass coverage type and doesn't consider other factors that may also influence the success rate, such as offensive formation, drop-back type, defenders in the box, and receiver route.

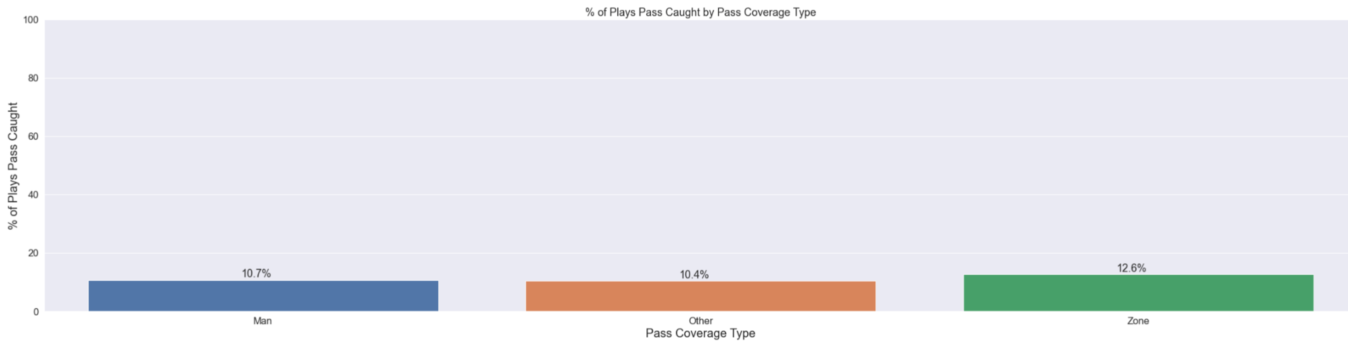

Fig 15. Percent of plays that lead to First Down/Score by Pass Coverage Type

Modeling:

After exploring trends in pass attempts in the NFL, I am now turning my attention to building a predictive model for the likelihood of a successful pass play. The goal is to develop a model that accurately predicts whether a pass play will result in a first down or touchdown, given various contextual features such as down, distance, and field position. To achieve this, I will use a machine learning approach, specifically classification algorithms, to analyze a large dataset of NFL plays and identify the key factors that contribute to a successful pass play. Through this modeling process, I am aiming to uncover important insights and develop a useful tool for coaches, players, and fans to understand better and predict the outcome of a pass play in the NFL.

The analysis splits the data into 3 sets: plays with 3, 4, and 5 eligible receivers, and I created separate models for each set. To create the models, I utilized three different machine learning algorithms: Logistic Regression, Random Forest Classifier, and K-Nearest Neighbors. By using multiple models, the strengths and weaknesses of each could be evaluated, and the most accurate model could be selected for each receiver set. Each algorithm was trained on a subset of the dataset and then tested on a separate validation set. I will present the results of each model's performance and compare their effectiveness in predicting successful pass plays. Additionally, we will discuss the significance and limitations of each model and provide insights into potential areas of improvement for future research.

In addition to creating three separate models for each eligible receiver set, several other steps were taken to prepare the data for modeling. First, a new variable called 'first_down_achieved' was created using the existing variables 'yardsToGo' and 'playResult'. This variable was used as the classification target for each play in the dataset.

Next, the data were merged into a single row based on the number of catch-eligible receivers, allowing analysis and modeling to be performed on each receiver set separately. To account for the categorical features in the dataset, dummy variables were created for the following: offensive formation, drop-back type, pass coverage, pass coverage type, defenders in the box, and receiver route run. These dummy variables were used as predictors in the models, along with the number of catch-eligible receivers on each play.

3-Receiver Sets:

Based on the given data for 3-eligible receiver sets[Fig.16], it is difficult to determine which model is the best since all three models have relatively low accuracy, precision, and recall. However, if we had to choose, we can eliminate the Random Forest model since it has a precision and recall of 0, indicating that it did not correctly identify any positive cases. Between the Logistic Regression and KNN models, the KNN model has a slightly higher precision and recall than the Logistic Regression model. Therefore, we can consider the KNN model to be the better choice for predicting whether a first down or touchdown will occur in 3-eligible receiver sets. However, it's important to note that the overall performance of these models is not ideal, and further exploration and feature engineering may be necessary to improve the model's accuracy.

```
Classification Report for Logistic Regression model - 3 Receivers
              precision    recall  f1-score   support

           0       0.58      0.97      0.72       112
           1       0.40      0.02      0.05        82

    accuracy                           0.57       194
   macro avg       0.49      0.50      0.39       194
weighted avg       0.50      0.57      0.44       194

Classification Report for Random Forest model - 3 Receivers
              precision    recall  f1-score   support

           0       0.58      1.00      0.73       112
           1       0.00      0.00      0.00        82

    accuracy                           0.58       194
   macro avg       0.29      0.50      0.37       194
weighted avg       0.33      0.58      0.42       194

Classification Report for KNN model - 3 Receivers
              precision    recall  f1-score   support

           0       0.59      0.90      0.71       112
           1       0.52      0.15      0.23        82

    accuracy                           0.58       194
   macro avg       0.56      0.52      0.47       194
weighted avg       0.56      0.58      0.51       194
```

Fig 16. Classification Reports for 3-WR models

4-Receiver Sets:

Based on the data for 4-eligible receiver sets[Fig.17], the best model is Random Forest. It has the highest accuracy at 0.62, and a precision score of 1.00 indicates that the model has correctly predicted all true positives, which is an important metric for predicting successful plays. However, the recall score of 0.01 is very low, which means the model may have missed a large number of actual successful plays. The Random Forest model still outperforms the other two models, Logistic Regression and KNN, in terms of accuracy and precision. But, the Logistic Regression model is only slightly behind on accuracy but has a much better recall, even though it's not a great recall.

```
Classification Report for Logistic Regression model - 4 Receivers
              precision    recall  f1-score   support

           0       0.63      0.92      0.75       314
           1       0.46      0.11      0.18       193

    accuracy                           0.61       507
   macro avg       0.54      0.52      0.46       507
weighted avg       0.56      0.61      0.53       507


Classification Report for Random Forest model - 4 Receivers
              precision    recall  f1-score   support

           0       0.62      1.00      0.77       314
           1       1.00      0.01      0.02       193

    accuracy                           0.62       507
   macro avg       0.81      0.51      0.39       507
weighted avg       0.77      0.62      0.48       507


Classification Report for KNN model - 4 Receivers
              precision    recall  f1-score   support

           0       0.61      0.91      0.73       314
           1       0.27      0.05      0.09       193

    accuracy                           0.59       507
   macro avg       0.44      0.48      0.41       507
weighted avg       0.48      0.59      0.49       507
```

Fig 17. Classification Reports for 4-WR models

5-Receiver Sets:

Based on the data below[Fig. 18], it is difficult to determine the best model for 5-eligible receiver sets. However, it appears that Random Forest has the highest precision score of 0.5, meaning it correctly identified 50% of the first downs achieved by the offense. The precision score for Logistic Regression is 0.00, indicating that the model did not correctly identify any of the first downs achieved by the offense. The KNN model has a precision score of 0.31, which is lower than the Random Forest model. However, the recall score for all three models is very low, indicating that they are not very effective at identifying all instances of first downs achieved by the offense. While the accuracy score is the same across all 3 models.

Classification Report for Logistic Regression model - 5 Receivers

	precision	recall	f1-score	support
0	0.67	1.00	0.80	1191
1	0.00	0.00	0.00	581
accuracy			0.67	1772
macro avg	0.34	0.50	0.40	1772
weighted avg	0.45	0.67	0.54	1772

Classification Report for Random Forest model - 5 Receivers

	precision	recall	f1-score	support
0	0.67	1.00	0.80	1191
1	0.50	0.01	0.02	581
accuracy			0.67	1772
macro avg	0.59	0.50	0.41	1772
weighted avg	0.62	0.67	0.55	1772

Classification Report for KNN model - 5 Receivers

	precision	recall	f1-score	support
0	0.67	0.99	0.80	1191
1	0.31	0.01	0.02	581
accuracy			0.67	1772
macro avg	0.49	0.50	0.41	1772
weighted avg	0.55	0.67	0.54	1772

Fig 18. Classification Reports for 5-WR models

Takeaways:

Based on the metrics provided, it seems that the Random Forest model performed the best overall for the 3 datasets. It had the highest accuracy and precision for the 4-receiver set, the highest accuracy and recall for the 3-receiver set, and the highest precision and recall for the 5-receiver set.

The Logistic Regression model had poor performance across all three datasets, with low precision and recall scores. The KNN model had the highest precision for the 3-receiver set, but overall its performance was also relatively poor compared to the Random Forest model.

Therefore, based on the provided data, it appears that the Random Forest model is the best model for predicting the success of passing plays with different eligible receiver sets. However, it is important to note that additional evaluation metrics and analysis may be needed to fully understand the strengths and weaknesses of each model.

1. https://www.sportsbusinessjournal.com/Journal/Issues/2023/01/09/Upfront/top-100-telecasts.aspx
2. https://www.forbes.com/legacy/forbes/2000/1211/6615132tab2_table.shtml
3. https://www.forbes.com/sites/mikeozanian/2022/10/27/nba-team-values-2022-for-the-first-time-in-two-decades-the-top-spot-goes-to-a-franchise-thats-not-the-knicks-or-lakers/?sh=33dea4121cce
4. https://www.kaggle.com/competitions/nfl-big-data-bowl-2023