

Tim Hulak  
IST-718 Big Data Analytics  
12/4/2021  
Prof. Jillian Lando

# SOAR Project Update

## Specification

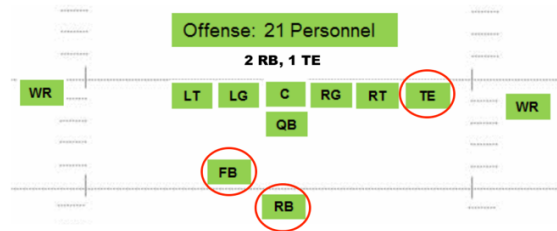
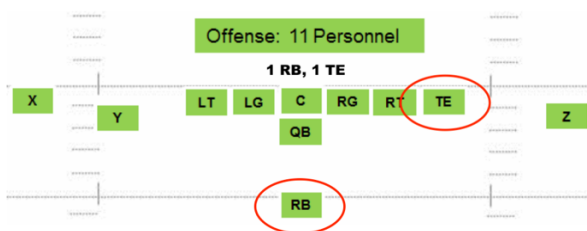
The primary problem that is trying to be solved is how can we use data to help the San Francisco 49ers football team improve their performance and appear in the playoffs. After some exploration of the data, the primary questions have morphed a little bit. The new questions are as follows:

- Which Offensive formations have seen the most success (measured in yards gained)?
- Which routes have seen the most success (measured in yards gained)?
- Which defensive formations have been the weakest (measured in yards gained against that defensive formation)?
- What are the attributes and characteristics of divisional opponents and how can they be exploited?
- 

## Observation

The 49ers generated 3,744 passing yards while deploying 6 offensive formations against 12 defensive formations. When analyzing the 49ers most successful passing formations, it appears that they have had the most success while in 11 Personnel (1,486 yards gained, 39.7% of total) and 21 Personnel (1,661 yards 44.4% of total). 11 Personnel seemed to perform best against a 4-2-5 defensive formation (sometimes called *Nickel* since there are 5 defensive backs) with 489 yards gained (13.1% of total yards) and 21 Personnel seemed to perform best against a 4-3-4 defensive formation (a form of *Base defense*, i.e., standard personnel) with 1003 yards (26.9% of total yards).

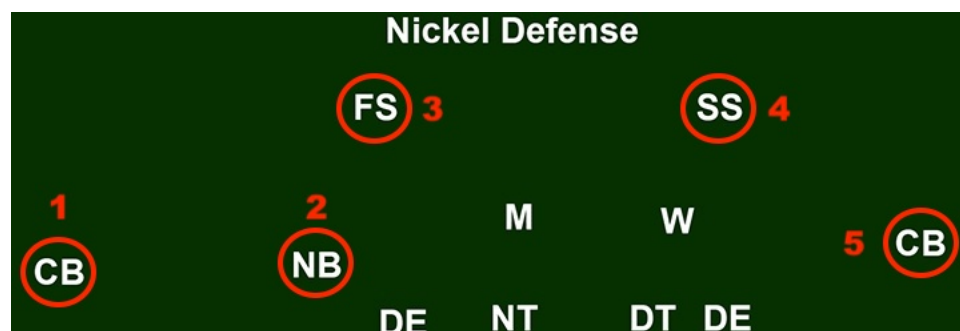
personnelO	0 DL, 4 LB, 7 DB	0 DL, 5 LB, 6 DB	2 DL, 3 LB, 6 DB	2 DL, 4 LB, 5 DB	3 DL, 1 LB, 7 DB	3 DL, 2 LB, 6 DB	3 DL, 3 LB, 5 DB	3 DL, 4 LB, 4 DB	4 DL, 1 LB, 6 DB	4 DL, 2 LB, 5 DB	4 DL, 3 LB, 4 DB	5 DL, 2 LB, 4 DB
0 RB, 1 TE, 4 WR	0	0	0	0	0	0	26	0	3	0	0	0
1 RB, 1 TE, 3 WR	12	9	138	61	34	175	255	4	280	489	29	0
1 RB, 2 TE, 2 WR	0	0	0	22	0	0	38	64	0	36	263	0
1 RB, 3 TE, 1 WR	0	0	0	11	0	0	0	0	0	0	0	0
2 RB, 1 TE, 2 WR	0	0	0	79	0	0	205	225	0	118	1003	31
2 RB, 2 TE, 1 WR	0	0	0	0	0	0	0	0	0	0	134	0



11 Personnel and 21 Personnel Packages reference diagrams

The Seattle Seahawks almost exclusively pass out of 11 Personnel and have the most success against Nickel defensive formations (formations with 5 defensive backs). Seattle generated 2,042 of their 2,751 (74.2%) yards out of 11 Personnel and 1,357 of their 2,751 (49.3%) yards against Nickel defensive formations while in 11 Personnel.

personnelO	1 DL, 3 LB, 7 DB	1 DL, 4 LB, 6 DB	2 DL, 3 LB, 6 DB	2 DL, 4 LB, 5 DB	3 DL, 2 LB, 6 DB	3 DL, 3 LB, 5 DB	3 DL, 4 LB, 4 DB	4 DL, 1 LB, 6 DB	4 DL, 2 LB, 5 DB	4 DL, 3 LB, 4 DB
0 RB, 1 TE, 4 WR	0	0	0	0	0	0	0	0	23	0
1 RB, 0 TE, 4 WR	0	0	15	0	5	0	0	0	0	0
1 RB, 1 TE, 3 WR	27	20	159	73	172	559	9	147	798	78
1 RB, 2 TE, 2 WR	0	0	0	17	0	0	18	0	10	236
2 RB, 0 TE, 3 WR	0	0	0	0	0	0	0	0	10	0
2 RB, 1 TE, 2 WR	0	0	0	0	0	0	9	0	18	48
6 OL, 1 RB, 0 TE, 3 WR	0	0	0	-2	0	0	45	9	2	38
6 OL, 1 RB, 1 TE, 2 WR	0	0	0	2	0	0	76	0	6	42
6 OL, 1 RB, 2 TE, 1 WR	0	0	0	0	0	0	5	0	0	25
6 OL, 2 RB, 0 TE, 2 WR	0	0	0	0	0	0	22	0	0	30



*Nickle defense reference (a defensive formation designed to better defend against pass plays)*

The Arizona Cardinals saw the most success while using 11 Personnel and were the most successful against the Nickel package, as well. They generated 2,030 out of their 2,687 yards while passing from 11 Personnel (75.5%) with 1,317 yards gained against the Nickel package (49.01%).

ARI Yards By Offensive Personnel vs Defensive Personnel														
personnelO	1 RB, 1 TE, 3 WR	60	21	117	246	297	365	16	110	667	81	39	11	0
	1 RB, 2 TE, 2 WR	0	0	0	0	4	4	27	0	18	340	0	29	0
	1 RB, 3 TE, 1 WR	0	0	0	0	0	0	0	17	0	79	0	0	35
	2 RB, 0 TE, 3 WR	0	0	0	0	11	3	0	0	0	0	0	0	0
	2 RB, 1 TE, 2 WR	0	0	0	0	0	6	8	0	13	46	0	13	0
	2 RB, 2 TE, 1 WR	0	0	0	0	0	0	0	0	0	4	0	0	0
	personnelD	1 DL, 3 LB, 7 DB	2 DL, 2 LB, 7 DB	2 DL, 3 LB, 6 DB	2 DL, 4 LB, 5 DB	3 DL, 2 LB, 6 DB	3 DL, 3 LB, 5 DB	3 DL, 4 LB, 4 DB	4 DL, 1 LB, 6 DB	4 DL, 2 LB, 5 DB	4 DL, 3 LB, 4 DB	5 DL, 1 LB, 5 DB	5 DL, 2 LB, 4 DB	6 DL, 1 LB, 4 DB

personnelO	personnelD	LA Yards
1 RB, 1 TE, 3 WR	2 DL, 2 LB, 7 DB	9
1 RB, 1 TE, 3 WR	2 DL, 3 LB, 6 DB	333
1 RB, 1 TE, 3 WR	2 DL, 4 LB, 5 DB	119
1 RB, 1 TE, 3 WR	3 DL, 2 LB, 6 DB	242
1 RB, 1 TE, 3 WR	3 DL, 3 LB, 5 DB	429
1 RB, 1 TE, 3 WR	4 DL, 1 LB, 6 DB	274
1 RB, 1 TE, 3 WR	4 DL, 2 LB, 5 DB	2626
1 RB, 1 TE, 3 WR	4 DL, 3 LB, 4 DB	109
1 RB, 1 TE, 3 WR	5 DL, 2 LB, 4 DB	29
1 RB, 2 TE, 2 WR	2 DL, 2 LB, 7 DB	0
1 RB, 2 TE, 2 WR	2 DL, 3 LB, 6 DB	0
1 RB, 2 TE, 2 WR	2 DL, 4 LB, 5 DB	0
1 RB, 2 TE, 2 WR	3 DL, 2 LB, 6 DB	0
1 RB, 2 TE, 2 WR	3 DL, 3 LB, 5 DB	0
1 RB, 2 TE, 2 WR	4 DL, 1 LB, 6 DB	0
1 RB, 2 TE, 2 WR	4 DL, 2 LB, 5 DB	104
1 RB, 2 TE, 2 WR	4 DL, 3 LB, 4 DB	0
1 RB, 2 TE, 2 WR	5 DL, 2 LB, 4 DB	5
1 RB, 3 TE, 1 WR	2 DL, 2 LB, 7 DB	0
1 RB, 3 TE, 1 WR	2 DL, 3 LB, 6 DB	0
1 RB, 3 TE, 1 WR	2 DL, 4 LB, 5 DB	0
1 RB, 3 TE, 1 WR	3 DL, 2 LB, 6 DB	0
1 RB, 3 TE, 1 WR	3 DL, 3 LB, 5 DB	0
1 RB, 3 TE, 1 WR	4 DL, 1 LB, 6 DB	0
1 RB, 3 TE, 1 WR	4 DL, 2 LB, 5 DB	0
1 RB, 3 TE, 1 WR	4 DL, 3 LB, 4 DB	31
1 RB, 3 TE, 1 WR	5 DL, 2 LB, 4 DB	0

## Analysis

A few models were used to try to predict yardage, which side of the field a play is likely to go to, which formations may be called during a particular game situation, and which route will be successful against a particular defensive formation. Linear Regression, Logistics regression, and multi-Logistic regression have been applied to the dataset. However, none of the models yielded very promising results (Logistic Regression even yielded a 19% accuracy rating). Further exploration into the options for modeling will be done.

## Recommendation

A deeper dive into the divisional opponents will be needed. This will include their offensive and defensive formation tendencies, success/weaknesses therein, and a deeper dive into the routes that wide receivers see the most success with.

In addition, there are many more granular options to consider for exploration. Things like focusing on individual quarter performance, down and distance performance, or completed

passes vs. incomplete passes broken down by defensive/offensive formation may appear to be getting into the weeds, but every inch counts on the football field.

The work that still needs to be done is on the data modeling front. Though multi-logistic regression has been attempted, it does not seem to be the best choice. Further exploration into the options for data modeling are needed in order to find the best model for this data and the questions being asked of it.