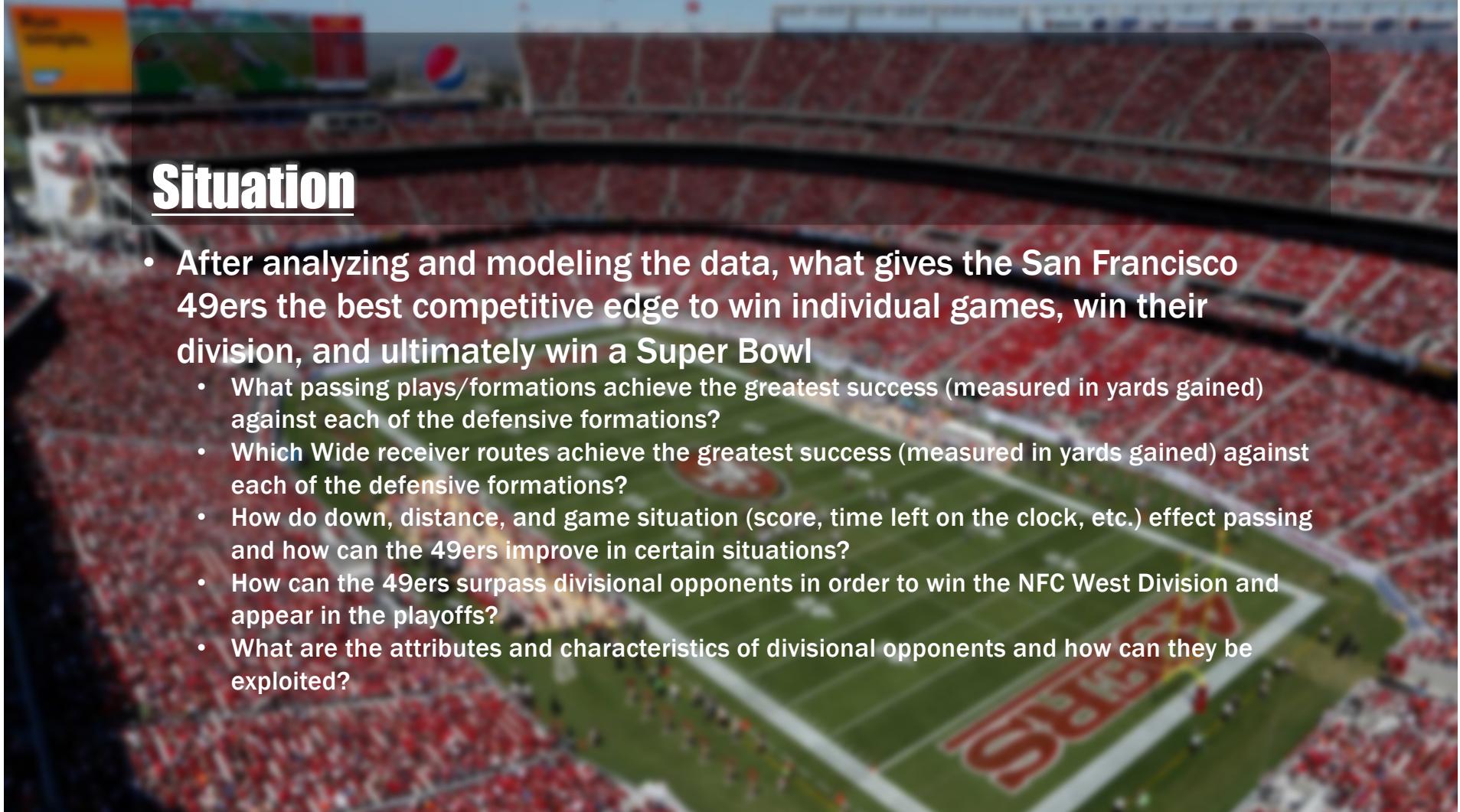


# [ Data ] Mining for Gold

Tim Hulak





## Situation

- After analyzing and modeling the data, what gives the San Francisco 49ers the best competitive edge to win individual games, win their division, and ultimately win a Super Bowl
  - What passing plays/formations achieve the greatest success (measured in yards gained) against each of the defensive formations?
  - Which Wide receiver routes achieve the greatest success (measured in yards gained) against each of the defensive formations?
  - How do down, distance, and game situation (score, time left on the clock, etc.) effect passing and how can the 49ers improve in certain situations?
  - How can the 49ers surpass divisional opponents in order to win the NFC West Division and appear in the playoffs?
  - What are the attributes and characteristics of divisional opponents and how can they be exploited?

## Situation: Background

-  vs.  (1976 - 2021): Games: 47, Record: 17-30 (36.2%)
-  vs.  (1950 - 2021): Games: 144, Record: 74-67-3 (51.3%)
-  vs.  (1951 - 2021): Games: 61, Record: 32-29 (52.5%)
-  Division: Games: 252, Record: 123-126-3 (48.8%)

## Raw Data Preview

quarter	down	yardsToGo	possessionTeam	yardlineNumber	offenseFormation	personnelO	defendersInTheBox	numberOfPassRushers	personnelD
0	1	15	ATL	20	I_FORM	2 RB, 1 TE, 2 WR	7.0	4.0	4 DL, 2 LB, 5 DB
8	1	10	ATL	39	SINGLEBACK	1 RB, 1 TE, 3 WR	7.0	4.0	4 DL, 2 LB, 5 DB
11	1	2	ATL	39	SHOTGUN	2 RB, 1 TE, 2 WR	6.0	4.0	4 DL, 2 LB, 5 DB
14	1	3	ATL	39	SHOTGUN	1 RB, 1 TE, 3 WR	6.0	5.0	4 DL, 1 LB, 6 DB
17	1	3	ATL	1	SHOTGUN	2 RB, 3 TE, 0 WR	8.0	6.0	6 DL, 3 LB, 2 DB

typeDropback	preSnapVisitorScore	preSnapHomeScore	gameClock	absoluteYardlineNumber	passResult	homeTeamAbbr	visitorTeamAbbr	week
TRADITIONAL	0.0	0.0	15:00:00	90.0	C	PHI	ATL	1
TRADITIONAL	0.0	0.0	13:10:00	49.0	I	PHI	ATL	1
TRADITIONAL	0.0	0.0	13:05:00	49.0	I	PHI	ATL	1
SCRAMBLE_ROLLOUT_LEFT	0.0	0.0	13:01:00	49.0	C	PHI	ATL	1
TRADITIONAL	0.0	0.0	10:59:00	11.0	I	PHI	ATL	1

playDirection	route	player	Situation	PlaySuccess	playResult
left	HITCH	J.Jones	1st & Long	Success	10
left	HITCH	C.Ridley	1st & 10	Failure	0
left	SCREEN	D.Freeman	2nd & Mid	Failure	0
left	CORNER	J.Jones	3rd & Mid	Success	33
left	OUT	D.Freeman	3rd & 1	Failure	0

## About the Data

- ✓ The original data came from the NFL Big Data Bowl Kaggle Competition (data from the 2018 NFL Season)
- ✓ Data is for passing plays only
  - ✓ Players table
  - ✓ Weeks Table (2018 Schedule)
  - ✓ 17 individual CSV files for plays in a given week
- ✓ After cleaning and merging, the data source for analysis had 16289 rows across 25 columns

## Data Preview

playDirection	route	player	Situation	PlaySuccess	playResult
left	HITCH	J.Jones	1st & Long	Success	10
left	HITCH	C.Ridley	1st & 10	Failure	0
left	SCREEN	D.Freeman	2nd & Mid	Failure	0
left	CORNER	J.Jones	3rd & Mid	Success	33
left	OUT	D.Freeman	3rd & 1	Failure	0

```

1 # Determine Success or Failure
2 conditions = []
3     (pass_caught['passResult'] == "I"),
4     # 1st Down Success
5     (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & 10') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
6     (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & 1') & (pass_caught['playResult'] > pass_caught['yardsToGo']),
7     (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & Forever') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.75),
8     (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & Long') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.60),
9     (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & Short') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
10    (pass_caught['down'] == 1) & (pass_caught['Situation'] == '1st & Mid') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
11    # 2nd Down Success
12    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & 10') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
13    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & 1') & (pass_caught['playResult'] > pass_caught['yardsToGo']),
14    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & Forever') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.75),
15    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & Long') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.60),
16    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & Short') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
17    (pass_caught['down'] == 2) & (pass_caught['Situation'] == '2nd & Mid') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
18    # 3rd Down Success
19    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & 10') & (pass_caught['playResult'] > pass_caught['yardsToGo']*80),
20    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & 1') & (pass_caught['playResult'] > pass_caught['yardsToGo']),
21    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & Forever') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.95),
22    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & Long') & (pass_caught['playResult'] > pass_caught['yardsToGo']*0.90),
23    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & Short') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
24    (pass_caught['down'] == 3) & (pass_caught['Situation'] == '3rd & Mid') & (pass_caught['playResult'] > pass_caught['yardsToGo']/2),
25    # 4th Down Success
26    (pass_caught['down'] == 4) & (pass_caught['playResult'] > pass_caught['yardsToGo'])]

```

## Defining Success or Failure

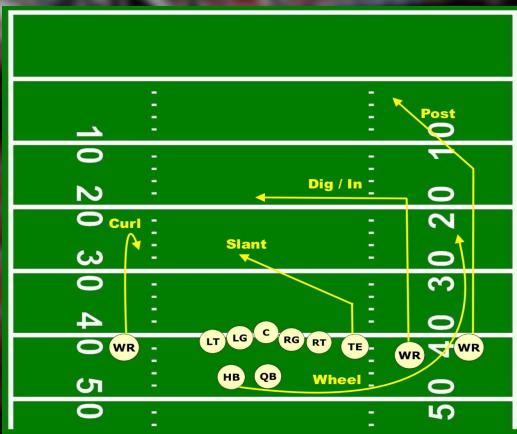
- Football is very granular and situational sport
  - 1st & 10 (standard)
  - Distance descriptions:
    - Short (1-3 yards)
    - Mid (4-7 yards)
    - Long (10-20 yards)
    - Forever (20+ yards)
- Situational Success
  - If pass is incomplete (not caught), automatic failure
  - Usually getting half the distance would be considered successful unless it is 4th down or there is a long way to get a 1st down (such as 3rd and 25); Success percentages based on how well a gain may set up the next play

## Some Context: Terminology

- Personnel: The mix of positions to make up 11 players for a team
- Formation: How players organize themselves on the field
- Offense: Team trying to score (throwing/running the ball)
- Defense: Team attempting to prevent the Offense from scoring
- Route: Predefined path a receiver will run and attempt to catch the ball (if thrown to them)
- Wide Receiver (WR): Offensive player tasked with running a route and catching a pass
- Tight End (TE): Offensive player who can block a defensive player or run a route to catch a pass
- Running Back (RB): Offensive player who can run the ball, block, or run the catch a pass
- Defensive Back (DB): Defensive player tasked with preventing a WR/TE from catching longer throws
- Linebacker (LB): Defensive player tasked with covering shorter throws
- Defensive Lineman (DL): Defensive player tasked with tackling the passer before he throws

# Some Context: Reference Diagrams

## Route Example



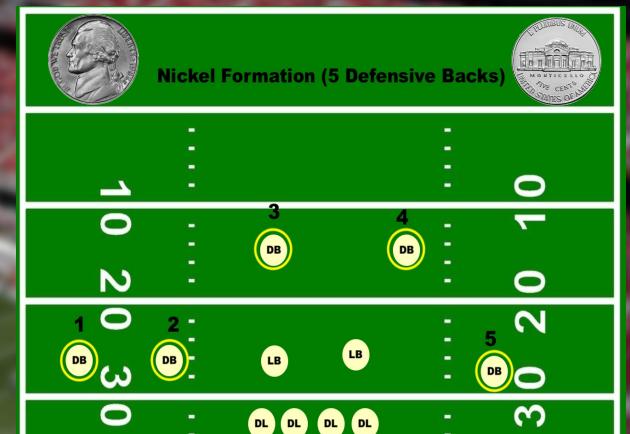
- Predefined paths for receivers to run
- Designed to exploit “holes” in the defense or pick on a “weaker” defender

## Offensive Personnel Example



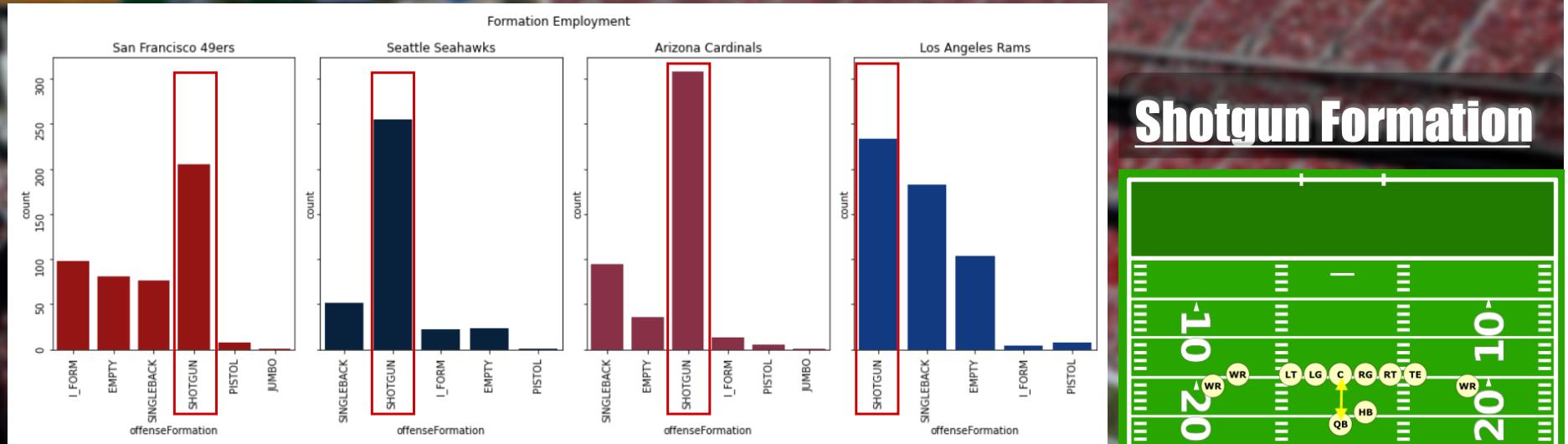
- Digit system for describing personnel
- Always equals 5
  - 2 RB, 1 TE (2 WR Implied)
  - 1 RB, 1 TE (3 WR Implied)
  - Lineman and QB are static:  $5 + 6 = 11$

## Defensive Personnel Example



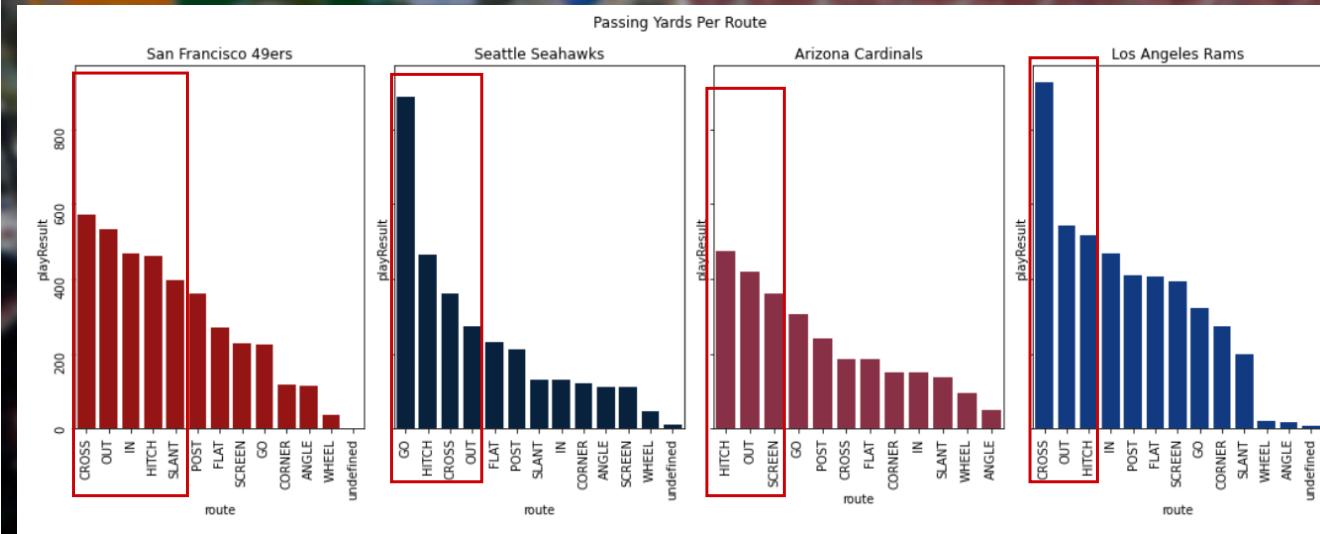
- Many Variations and personnel
- DL, LB, and DB can have numerous iterations
- Nickel (5 DBs) and Dime (6 DBs) [one greater than a Nickel] are typically deployed to stop a pass

# Observations & EDA

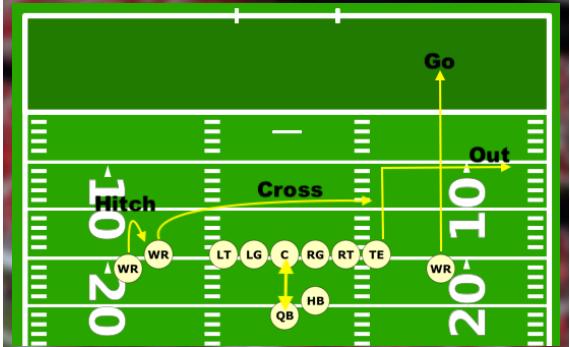


- Each team in the NFC West appears to favor the Shotgun Formation
- Since this is passing data, this makes sense because the Shotgun formation is a typical passing formation
- The Shotgun offers a few advantages such as:
  - More room for the Offensive Lineman to block
  - More room for the QB to maneuver
  - More time for a QB to make a decision

# Observations & EDA

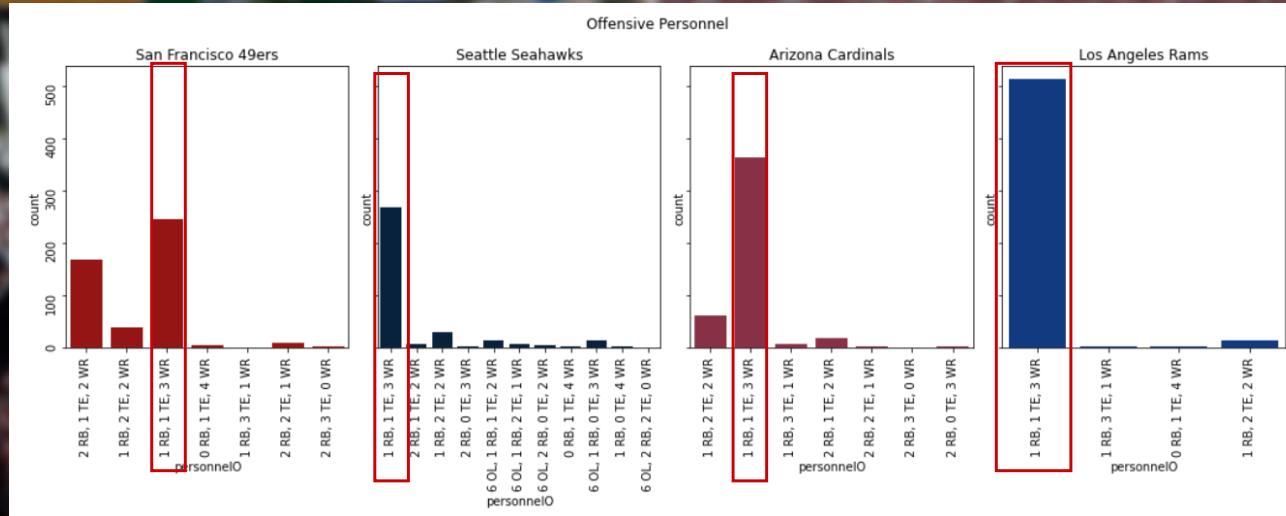


## Play Art

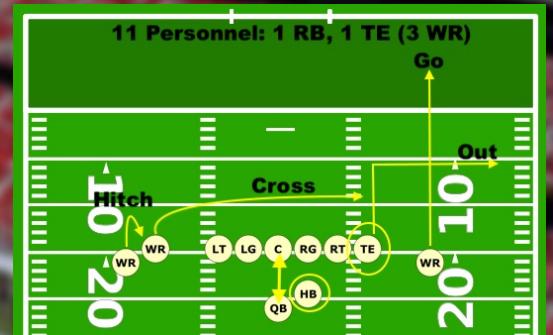


- There is some overlap across the NFC West when it comes to the Out Route
- San Francisco seems to have a healthy mix of short routes and routes designed to move defenders out of position
- Seattle tends to gain more yards using the Go route (likely due to speedy receivers)
- Los Angeles seems to gain more yards on Crossing routes (likely due to responsibility exchange and inside leverage)
- Arizona seems to gain more yards on shorter routes (get ball into receivers hands quickly before defense can react)

# Observations & EDA



## 11 Personnel



- Each of the teams in the NFC West seem to deploy 11 Personnel far more often than the other personnel packages
- Since they are historically a run-first team, San Francisco deploys 21, 11, and 12 personnel most often
- Los Angeles has the least diverse approach to passing and SEA appears to have the most diverse personnel package deployment
  - SF: 7 packages
  - SEA: 11 packages (most successful team in division, so maybe confusing defenses with seemingly endless personnel groupings is a key to their success)
  - ARI: 7 packages
  - LA: 4 packages

# Observations & EDA: Offense vs. Defense

LA Yards By Offensive Personnel vs Defensive Personnel											
personnelID	2 DL, 2 LB, 7 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	0	0	0	0	0
1 RB, 1 TE, 3 WR -	9	342	119	0	249	451	0	274	2784	109	29
1 RB, 2 TE, 2 WR -	0	0	0	0	0	0	0	0	106	0	5
1 RB, 3 TE, 1 WR	0	0	0	0	0	0	0	0	0	31	0

ARI Yards By Offensive Personnel vs Defensive Personnel											
personnelID	1 DL, 3 LB, 7 DB -	1 DL, 4 LB, 6 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 -
1 RB, 1 TE, 3 WR -	60	0	21	117	246	301	365	16	110	0	683
1 RB, 2 TE, 2 WR -	0	0	0	0	0	4	4	27	0	0	375
1 RB, 3 TE, 1 WR -	0	0	0	0	0	0	0	17	0	0	79
2 RB, 0 TE, 3 WR -	0	0	0	0	0	11	3	0	0	0	0
2 RB, 1 TE, 2 WR -	0	0	0	0	0	0	6	8	0	0	46
2 RB, 2 TE, 1 WR -	0	0	0	0	0	0	0	0	0	0	13
2 RB, 3 TE, 0 WR -	0	0	0	0	0	0	0	0	0	0	0

- LA saw success with the 11 personnel package against Nickel defense
- Other formations far less successful

- ARI saw great success with 11 personnel against numerous defensive formations, especially against Nickel and Dime

# Observations & EDA: Offense vs. Defense

SF Yards By Offensive Personnel vs Defensive Personnel													
personnelID	0 DL, 4 LB, 7 DB -	0 DL, 5 LB, 6 DB -	1 DL, 4 LB, 6 DB -	2 DL, 3 LB, 6 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 -	5 DL, 3 LB, 2 DB, 1 OL -
personnelID	0	0	0	0	0	0	26	0	13	0	0	0	0
0 RB, 1 TE, 4 WR -	0	0	0	0	0	0	26	0	13	0	0	0	0
1 RB, 1 TE, 3 WR -	12	9	3	149	61	34	175	255	4	280	494	29	0
1 RB, 2 TE, 2 WR -	0	0	0	0	22	0	0	38	64	0	36	263	1
1 RB, 3 TE, 1 WR -	0	0	0	0	11	0	0	0	0	0	0	0	0
2 RB, 1 TE, 2 WR -	0	0	0	0	79	0	0	205	225	0	118	1005	31
2 RB, 2 TE, 1 WR -	0	0	0	0	0	0	0	0	0	0	143	0	0
2 RB, 3 TE, 0 WR -	0	0	0	0	0	0	0	0	0	0	0	0	0

SEA Yards By Offensive Personnel vs Defensive Personnel													
personnelID	0 RB, 1 TE, 4 WR -	0 RB, 0 TE, 4 WR -	1 RB, 1 TE, 3 WR -	1 RB, 2 TE, 2 WR -	2 RB, 0 TE, 3 WR -	2 RB, 1 TE, 2 WR -	6 OL, 1 RB, 0 TE, 3 WR -	6 OL, 1 RB, 1 TE, 2 WR -	6 OL, 1 RB, 2 TE, 1 WR -	6 OL, 2 RB, 0 TE, 2 WR -	6 OL, 2 RB, 2 TE, 0 WR -	1 DL, 3 LB, 7 DB -	1 DL, 4 LB, 6 DB -
personnelID	0	0	0	0	0	0	0	0	0	0	0	0	0
0 RB, 1 TE, 4 WR -	0	0	0	0	0	0	0	12	0	0	23	0	0
1 RB, 0 TE, 4 WR -	0	0	15	0	5	0	0	0	0	0	0	0	0
1 RB, 1 TE, 3 WR -	27	20	159	130	172	628	9	153	914	78	0	0	0
1 RB, 2 TE, 2 WR -	0	0	0	17	0	0	41	0	10	236	0	0	0
2 RB, 0 TE, 3 WR -	0	0	0	0	0	0	0	0	10	0	0	0	0
2 RB, 1 TE, 2 WR -	0	0	0	0	0	0	0	9	0	18	48	0	0
6 OL, 1 RB, 0 TE, 3 WR -	0	0	0	-2	0	0	0	45	19	3	38	0	0
6 OL, 1 RB, 1 TE, 2 WR -	0	0	0	2	0	0	0	115	0	6	42	0	0
6 OL, 1 RB, 2 TE, 1 WR -	0	0	0	0	0	0	0	5	0	0	25	0	0
6 OL, 2 RB, 0 TE, 2 WR -	0	0	0	0	0	0	0	22	0	0	30	0	0
6 OL, 2 RB, 2 TE, 0 WR -	0	0	0	0	0	0	0	0	0	0	2	0	0

- SF was very successful when passing out of 11 and 21 personnel, specifically against Nickel and Base defense
  - Defense will typically deploy Nickel/Dime against pass and a Base formation to stop a run

- SEA saw success out of 11 Personnel against Nickel and Dime defensive formations

# Observations & EDA: Correlation

	quarter	yardlineNumber	offenseFormation	personnelO	defendersInTheBox	numberOfPassRushers	personnelD	typeDropback	absoluteYardlineNumber	passResult	playDirection	route	Situation	PlaySuccess	playResult
quarter	1	0.0036	0.048	-0.053	0.074	-0.0088	0.016	-0.0069	0.0066	0.01	0.0028	0.0083	0.051	0.0084	0.0033
yardlineNumber	0.0036	1	-0.0081	-0.017	-0.0098	-0.007	-0.011	-0.003	0.65	0.00016	-0.0015	0.017	0.0084	-0.021	0.011
offenseFormation	0.048	-0.0081	1	0.044	0.29	-0.0007	0.024	-0.12	0.002	0.0091	-0.0077	-0.0051	0.12	0.038	-0.06
personnelO	-0.053	-0.017	0.044	1	0.061	0.063	0.066	0.044	0.0031	-0.0016	-0.0019	-0.0057	0.023	0.00066	-0.0058
defendersInTheBox	0.074	-0.0098	0.29	0.061	1	0.009	0.2	-0.041	-0.0035	0.0071	0.0056	-0.0096	0.13	0.038	-0.023
numberOfPassRushers	-0.0088	-0.007	-0.0007	0.063	0.009	1	0.11	0.033	0.0042	-0.0043	-0.011	0.0038	0.049	0.0076	-0.0071
personnelD	0.016	-0.011	0.024	0.066	0.2	0.11	1	0.021	-0.0058	0.029	-0.0095	0.0093	0.044	0.022	-0.0015
typeDropback	-0.0069	-0.003	-0.12	0.044	-0.041	0.033	0.021	1	0.0064	0.039	0.0038	-0.0028	0.022	0.017	-0.0025
absoluteYardlineNumber	0.0066	0.65	0.002	0.0031	-0.0035	0.0042	-0.0058	0.0064	1	0.016	0.0033	0.017	0.057	-0.0012	-0.003
passResult	0.01	0.00016	0.0091	-0.0016	0.0071	-0.0043	0.029	0.039	0.016	1	-0.00075	0.14	0.021	0.67	-0.54
playDirection	0.0028	-0.0015	-0.0077	-0.0019	0.0056	-0.011	-0.0095	0.0038	0.0033	0.00075	1	-0.01	-0.0037	0.00019	-0.007
route	0.0083	0.017	-0.0051	-0.0057	-0.0096	0.0038	0.0093	-0.0028	0.017	0.14	-0.01	1	0.011	0.03	0.022
Situation	0.051	0.0084	0.12	0.023	0.13	0.049	0.044	0.022	0.057	0.021	-0.0037	0.011	1	0.064	-0.04
PlaySuccess	0.0084	-0.021	0.038	0.00066	0.038	0.0076	0.022	0.017	-0.0012	0.67	0.00019	0.03	0.064	1	-0.51
playResult	0.0033	0.011	-0.06	-0.0058	-0.023	-0.0071	-0.0015	-0.0025	-0.003	-0.54	-0.007	0.022	-0.04	-0.51	1

- We see that the highest correlations to Play Success are the pass result at 0.67(success or failure) and the play result at -0.54 (the yardage gained on a pass).
- This makes sense and verifies the importance of this analysis and modeling
- However, defensive variables (such as defensive formations and players in the box) seem to have little correlation.
- May suggest that offensive talent or scheme may be more important than facing a particular defensive formation generally

## Observations & EDA



- After EDA and correlation, the focus of this analysis has shifted from looking at formations/personnel to looking at routes and their probability of success.
- This is because, as seen in the picture above, teams will tailor their defense to offensive personnel (before the ball is even thrown). This means that there is a bias or preconceived plan to attack certain offensive personnel.
- **New Goal: Construct a Probability Based Situational Play Card**

## Analysis and Modeling

- Logistic Regression
- Situational Probability Calculations for Routes

# Logistic Regression

- Logistic Regression is a classification algorithm used to predict a categorical variable.
- In this model, we attempt to predict the ‘Success’ or ‘Failure’ of a pass play in a given situation
- The first step is to encode the categorical variables and scale them to be fit in the model

quarter	yardlineNumber	defendersInTheBox	numberOfPassRushers	absoluteYardlineNumber	passResult	playDirection	route	Situation	PlaySuccess	playResult	
0	1	20	7.0	4.0	90.0	C	left	HITCH	1st & Long	Success	10
8	1	39	7.0	4.0	49.0	I	left	HITCH	1st & 10	Failure	0
11	1	39	6.0	4.0	49.0	I	left	SCREEN	2nd & Mid	Failure	0
14	1	39	6.0	5.0	49.0	C	left	CORNER	3rd & Mid	Success	33
17	1	1	8.0	6.0	11.0	I	left	OUT	3rd & 1	Failure	0



quarter	yardlineNumber	defendersInTheBox	numberOfPassRushers	absoluteYardlineNumber	passResult	playDirection	route	Situation	PlaySuccess	playResult	
0	1	20	7.0	4.0	90.0	0	0	5	3	1	10
8	1	39	7.0	4.0	49.0	1	0	5	1	0	0
11	1	39	6.0	4.0	49.0	1	0	9	9	0	0
14	1	39	6.0	5.0	49.0	0	0	1	14	1	33
17	1	1	8.0	6.0	11.0	1	0	7	11	0	0

# Logistic Regression: Coefficients

	Var	Coef
0	quarter	-0.001329
1	yardlineNumber	-0.036236
2	defendersInTheBox	0.532127
3	numberOfPassRushers	-0.008231
4	absoluteYardlineNumber	0.035646
5	passResult	-2.900996
6	playDirection	-0.025962
7	route	0.063148
8	Situation	0.345892
9	playResult	5.578918

- Negative coefficients indicate that odds of belonging to the ‘Success’ feature decrease
  - Quarter increases (as game goes on longer) , odds of success *decrease*
  - Yard Line is further away, odds of success *decrease*
  - Number of Pass Rushers increases, odds of success *decrease*
- Positive coefficients indicate that odds of belonging to the ‘Success’ feature increase
  - The fewer Defenders in the Box, odds of success *increase*
  - The closer to the goal relative to the entire field (Absolute Yard Line : field measured from 0 -100) , odds of success *increase*
  - The more Yards gained (Play Result) , odds of success *increase*

# Logistic Regression: Model Performance

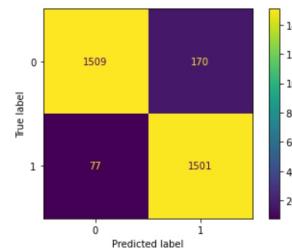
- The Model performed with 92.4% accuracy, meaning that, given the data, we can accurately predict the success or failure of a play around 92% of the time
- The confusion matrix shows that 1,509 plays were successfully predicted as a failure and 1,501 plays were successfully predicted as a success.
- 247 plays were incorrectly predicted and 3,010 plays were correctly predicted
- The classification report shows a precision of 95% correctly classified for failure and 90% correctly classified for success. Both show a harmonic mean of 92% accurate classification.

```
In [84]: 1 # Accuracy  
2 round(accuracy_score(y_test, y_pred),3)
```

```
Out[84]: 0.924
```

```
In [89]: 1 plot_confusion_matrix(log_model, scaled_X_test, y_test)
```

```
Out[89]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7ff31ad30fa0>
```



	precision	recall	f1-score	support
0	0.95	0.90	0.92	1679
1	0.90	0.95	0.92	1578
accuracy			0.92	3257
macro avg	0.92	0.92	0.92	3257
weighted avg	0.93	0.92	0.92	3257



## Recommendation

### **Routes (12)**

- HITCH
- SCREEN
- CORNER
- OUT
- FLAT
- IN
- POST
- CROSS
- ANGLE
- GO
- SLANT
- WHEEL

### **Situations (21)**

- 1st & 1
- 1st & 10
- 1st & Forever
- 1st & Long
- 1st & Mid
- 1st & Short
- 2nd & 1
- 2nd & Forever
- 2nd & Long
- 2nd & Mid
- 2nd & Short
- 3rd & 1
- 3rd & Forever
- 3rd & Long
- 3rd & Mid
- 3rd & Short
- 4th & 1
- 4th & Forever
- 4th & Long
- 4th & Mid
- 4th & Short

# Recommendation: Play Card



- A “Play Card” is a laminated page jam packed with plays for passing, running, and situations.
  - The card is constructed each week as a result of studying an upcoming opponent
  - Quarterback may have a condensed version on a wristband

# Recommendation: Play Card

Probabilities to construct play card

```
In [55]: 1 situation = []
2 route = []
3 prob = []
4
5
6 for i in (pass_caught.route.unique()):
7     for j in sorted(list(pass_caught.Situation.unique())):
8         situation.append(j)
9         route.append(i)
10    prob.append((round((pass_caught.loc[(pass_caught['Situation'] == j) & (pass_caught['PlaySuccess'] == 'Success')] (pass_caught[t['route']] == i), 'playResult').sum())/(pass_caught.loc[pass_caught['Situation'] == j, 'playResult'].sum()), 2))
11
12 card = pd.DataFrame({
13     "Situation": situation,
14     "Route": route,
15     "Probability": prob
16 })
17 card = card[card['Route'] != 'undefined']

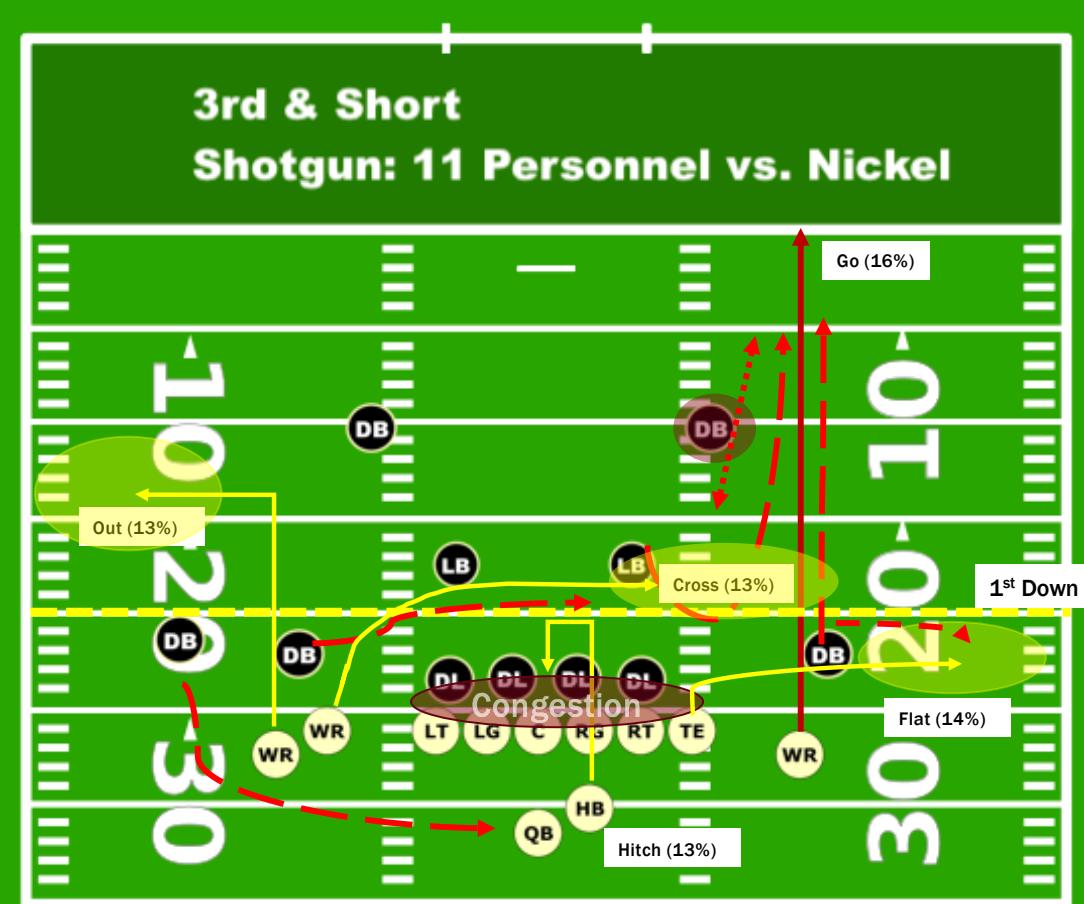
In [56]: 1 card.groupby(['Situation','Route']).max().sort_values(['Situation','Probability'], ascending=False)
```

Game # 1: Seattle Seahawks													
1st Down			2nd Down			3rd Down			4th Down				
Situation	Route/Progression	Probability	Situation	Route/Progression	Probability	Situation	Route/Progression	Probability	Situation	Route/Progression	Probability		
1st & 10	CROSS	0.15	2nd & 1	GO	0.21	3rd & 1	CROSS	0.26	4th & 1	FLAT	0.26		
	GO	0.13		OUT	0.15		FLAT	0.17		OUT	0.16		
	HITCH	0.11		HITCH	0.11		OUT	0.12		CROSS	0.15		
	OUT	0.11		2nd & Short			3rd & Short			HITCH	0.13		
1st & 1	CROSS	0.08		HITCH	0.13		GO	0.16	4th & Short	FLAT	0.27		
1st & Short	FLAT	0.22		FLAT	0.11		FLAT	0.14		CORNER	0.17		
	CROSS	0.13		IN	0.10		CROSS	0.13		WHEEL	0.12		
	OUT	0.13	2nd & Mid	HITCH	0.11		OUT	0.13		GO	0.1		
	SLANT	0.13		GO	0.12	3rd & Mid	OUT	0.16	4th & Mid	GO	0.32		
1st & Mid	CROSS	0.20		CROSS	0.12		HITCH	0.14		IN	0.18		
	OUT	0.14		FLAT	0.12		GO	0.13		HITCH	0.11		
	FLAT	0.13	2nd & Long	HITCH	0.13		CROSS	0.09		SLANT	0.08		
	SCREEN	0.11		GO	0.12	3rd & Long	GO	0.13	4th & Long	IN	0.28		
1st & Long	POST	0.11		CROSS	0.09		HITCH	0.08		OUT	0.26		
	IN	0.10		SCREEN	0.08		IN	0.08		GO	0.17		
	HITCH	0.10	2nd & Forever	GO	0.21		3rd & Forever			PUNT	NA		
1st & Forever	SCREEN	0.32		HITCH	0.08		POST			4th & Forever			
	GO	0.13		OUT	0.07		PUNT			NA			
	OUT	0.11		NA			NA			NA			
	POST	0.11		NA			NA			NA			

## Recommendation: Play Art Example

3rd & Short	GO	0.16
	FLAT	0.14
	CROSS	0.13
	HITCH	0.13
	OUT	0.13

1. Go: Deep pass/clear out
2. Flat: If Go clears/DB trails
3. Cross: If Go clears LB Flat Clears DB
4. Hitch: If LBs trail, hit middle
5. Out: Clear DB for cross/ hit if DB blitz
6. All of this goes through a QB's head before play and during the play (within 3-4 seconds before the ball is thrown)





# [ Data ] Mining for Gold

Tim Hulak

Questions?

