

NFL Performance Data

Jordan Hyatt, Patrick Bulger,
Timothy Hulak, Cal Wardell



Raw Data Preview

```
In [4]: NFL.head(10)
```

```
Out[4]:
```

	play_id	game_id	home_team	away_team	posteam	posteam_type	defteam	side_of_field	yardline_100	game_date	quarter_seconds_remaining	half_secs
0	46	2009091000	PIT	TEN	PIT	home	TEN	TEN	30.0	2009-09-10	900.0	
1	68	2009091000	PIT	TEN	PIT	home	TEN	PIT	58.0	2009-09-10	893.0	
2	92	2009091000	PIT	TEN	PIT	home	TEN	PIT	53.0	2009-09-10	856.0	
3	113	2009091000	PIT	TEN	PIT	home	TEN	PIT	56.0	2009-09-10	815.0	
4	139	2009091000	PIT	TEN	PIT	home	TEN	PIT	56.0	2009-09-10	807.0	
5	162	2009091000	PIT	TEN	TEN	away	PIT	TEN	98.0	2009-09-10	796.0	
6	183	2009091000	PIT	TEN	TEN	away	PIT	TEN	98.0	2009-09-10	760.0	
7	207	2009091000	PIT	TEN	TEN	away	PIT	TEN	94.0	2009-09-10	731.0	
8	228	2009091000	PIT	TEN	TEN	away	PIT	TEN	96.0	2009-09-10	694.0	
9	253	2009091000	PIT	TEN	PIT	home	TEN	TEN	43.0	2009-09-10	684.0	

```
In [5]: NFL.shape
```

```
Out[5]: (449371, 255)
```

About the Data

- ✓ The dataset contains 449,371 observations across 255 variables.
- ✓ Each variable represents information about a given play during an NFL football game.
- ✓ Each record represents a specific play during a given game.
- ✓ The data contains information on plays from the 2009 - 2018 seasons.

Raw Data Preview

```
In [9]: col_del_list = [
    "no_score_prob", "opp_fg_prob", "opp_safety_prob", "opp_td_prob", "fg_prob",
    "safety_prob", "td_prob", "extra_point_prob", "two_point_conversion_prob",
    "ep", "epa", "total_home_epa", "total_away_epa", "total_home_rush_epa",
    "total_away_rush_epa", "total_home_pass_epa", "air_epa", "yac_epa", "comp_air_epa",
    "comp_yac_epa", "total_home_comp_air_epa", "total_away_comp_air_epa",
    "total_home_comp_yac_epa", "total_away_comp_yac_epa", "total_home_raw_air_epa",
    "total_away_raw_air_epa", "total_home_raw_yac_epa", "total_away_raw_yac_epa",
    "wp", "def_wp", "home_wp", "away_wp", "wpa", "home_wp_post", "away_wp_post",
    "total_home_rush_wpa", "total_away_rush_wpa", "total_home_pass_wpa",
    "total_away_pass_wpa", "air_wpa", "yac_wpa", "comp_air_wpa", "comp_yac_wpa",
    "total_home_comp_air_wpa", "total_away_comp_air_wpa", "total_home_comp_yac_wpa",
    "total_home_raw_air_wpa", "total_away_raw_air_wpa", "total_home_raw_yac_wpa",
    "total_away_raw_yac_wpa", "passer_player_id", "receiver_player_id",
    "rusher_player_id", "lateral_receiver_player_id", "lateral_rusher_player_id",
    "lateral_sack_player_id", "interception_player_id",
    "lateral_interception_player_id", "punt_returner_player_id",
    "lateral_punt_returner_player_id", "kickoff_returner_player_id",
    "lateral_kickoff_returner_player_id", "punter_player_id", "kicker_player_id",
    "own_kickoff_recovery_player_id", "blocked_player_id",
    "tackle_for_loss_1_player_id", "tackle_for_loss_2_player_id",
    "qb_hit_1_player_id", "qb_hit_2_player_id", "forced_fumble_player_1_player_id",
    "forced_fumble_player_2_player_id",
    "solo_tackle_1_player_id", "solo_tackle_2_player_id", "assist_tackle_1_player_id",
    "assist_tackle_2_player_id", "assist_tackle_3_player_id",
    "assist_tackle_4_player_id", "pass_defense_1_player_id", "pass_defense_2_player_id",
    "fumbled_1_player_id", "fumbled_2_player_id", "fumble_recovery_1_team",
    "fumble_recovery_1_player_id", "fumble_recovery_2_player_id", "penalty_player_id",
    "total_away_pass_epa"
]

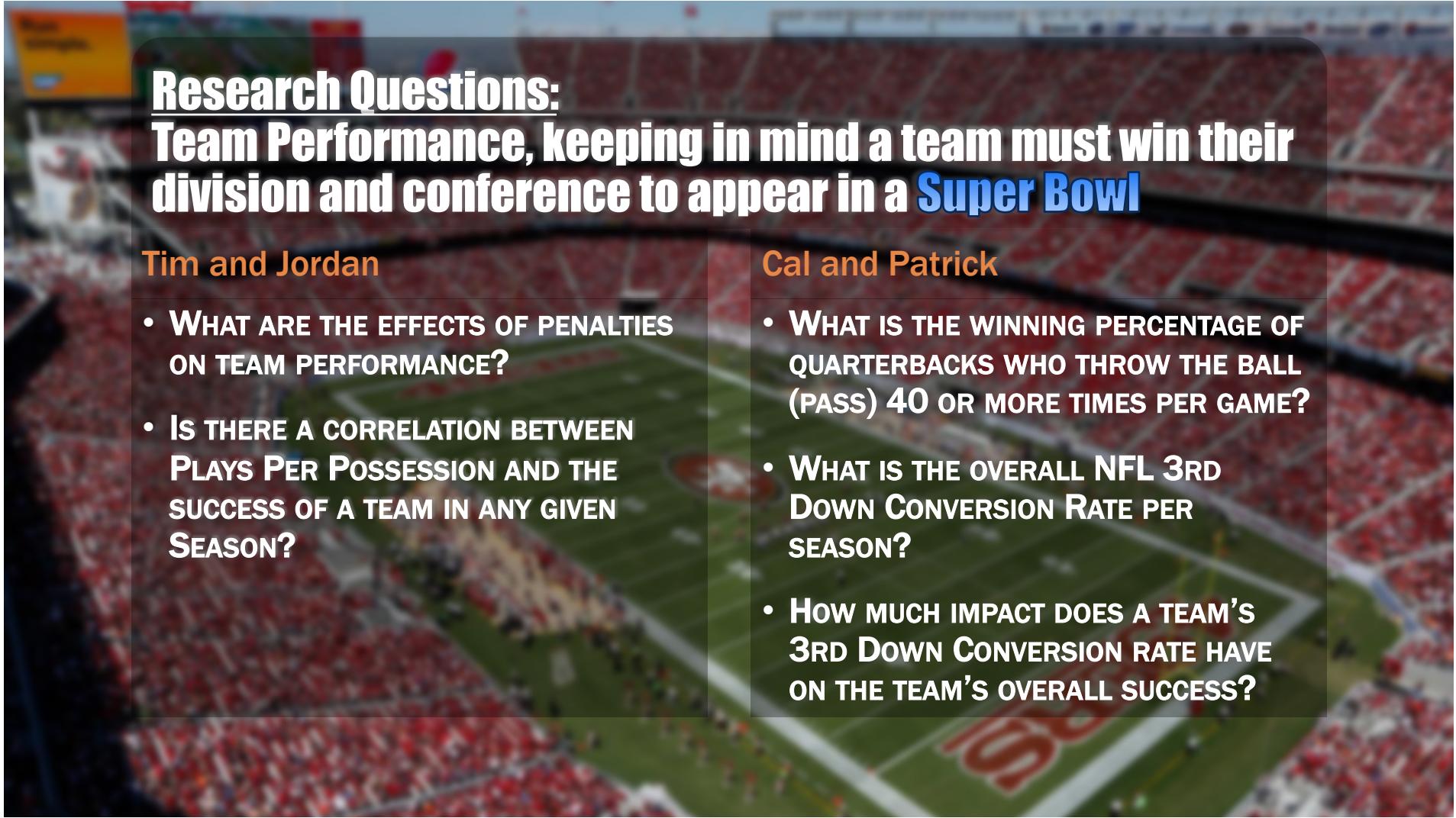
for column in col_del_list:
    del NFL[column]

In [10]: len(col_del_list)
Out[10]: 87

In [11]: NFL.replace("JAC", "JAX", inplace= True)
NFL.replace("SD", "LAC", inplace= True)
NFL.replace("STL", "LA", inplace= True)
```

Initial Data Cleaning

- Given the large number of variables, many of them can be discarded. Specifically, the statistical columns for probability after a given play.
- There are many fields in the data that represent the probability of a given action occurring (such as "fg_prob", which represents the probability the team will kick a field goal on that given play).
- While these advanced metrics are interesting, they are not necessary for the analysis and may be removed.
- This will leave the data with 168 of the 255 columns. 3 teams also changed their team abbreviation during the time span of the data.
- Since they are the same team, just have a different abbreviation, the old abbreviations were replaced with the new abbreviations.



Research Questions:

Team Performance, keeping in mind a team must win their division and conference to appear in a Super Bowl

Tim and Jordan

- **WHAT ARE THE EFFECTS OF PENALTIES ON TEAM PERFORMANCE?**
- **IS THERE A CORRELATION BETWEEN PLAYS PER POSSESSION AND THE SUCCESS OF A TEAM IN ANY GIVEN SEASON?**

Cal and Patrick

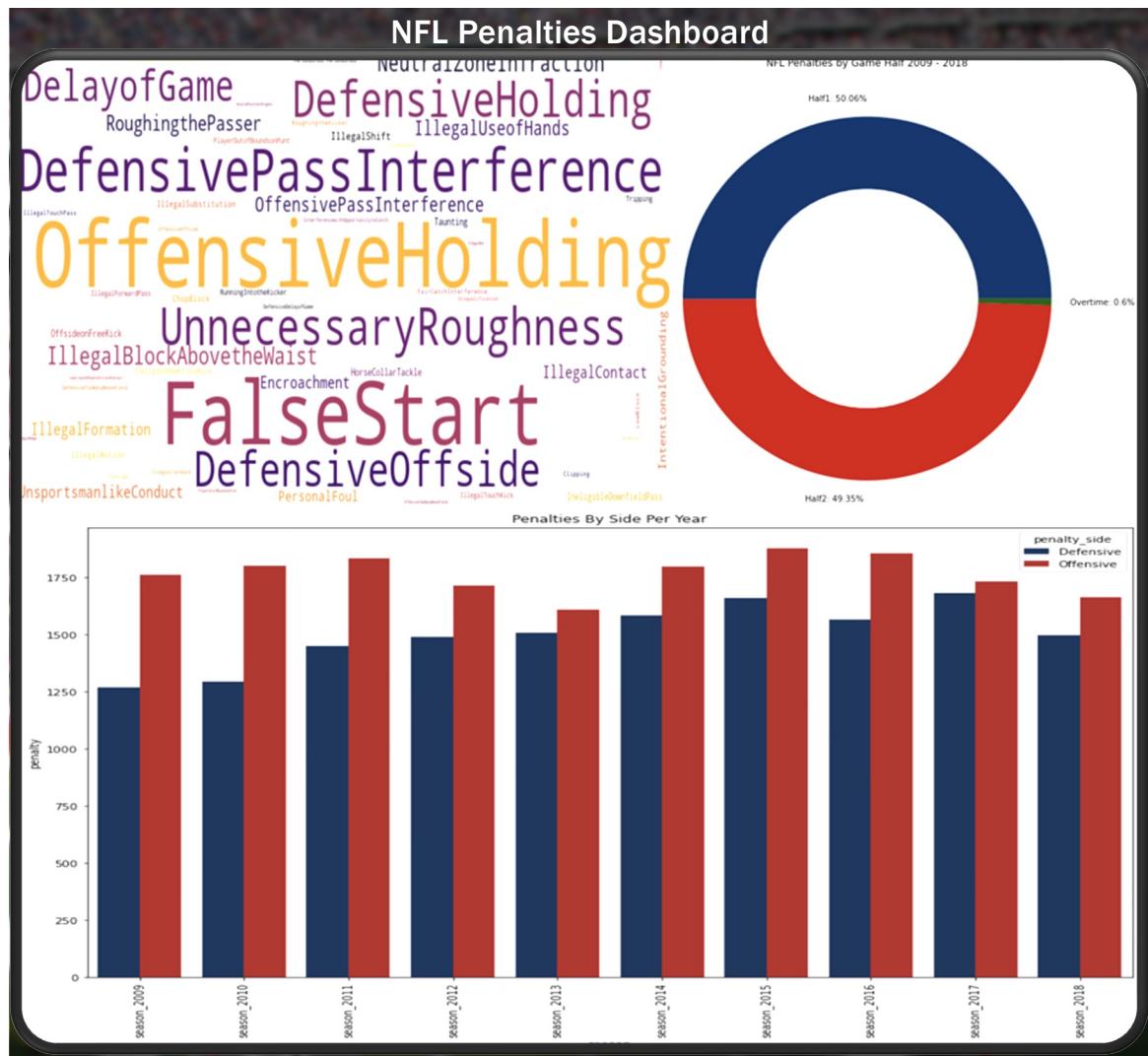
- **WHAT IS THE WINNING PERCENTAGE OF QUARTERBACKS WHO THROW THE BALL (PASS) 40 OR MORE TIMES PER GAME?**
- **WHAT IS THE OVERALL NFL 3RD DOWN CONVERSION RATE PER SEASON?**
- **HOW MUCH IMPACT DOES A TEAM'S 3RD DOWN CONVERSION RATE HAVE ON THE TEAM'S OVERALL SUCCESS?**

WHAT ARE THE EFFECTS OF PENALTIES ON TEAM PERFORMANCE?

Key Program Feature:

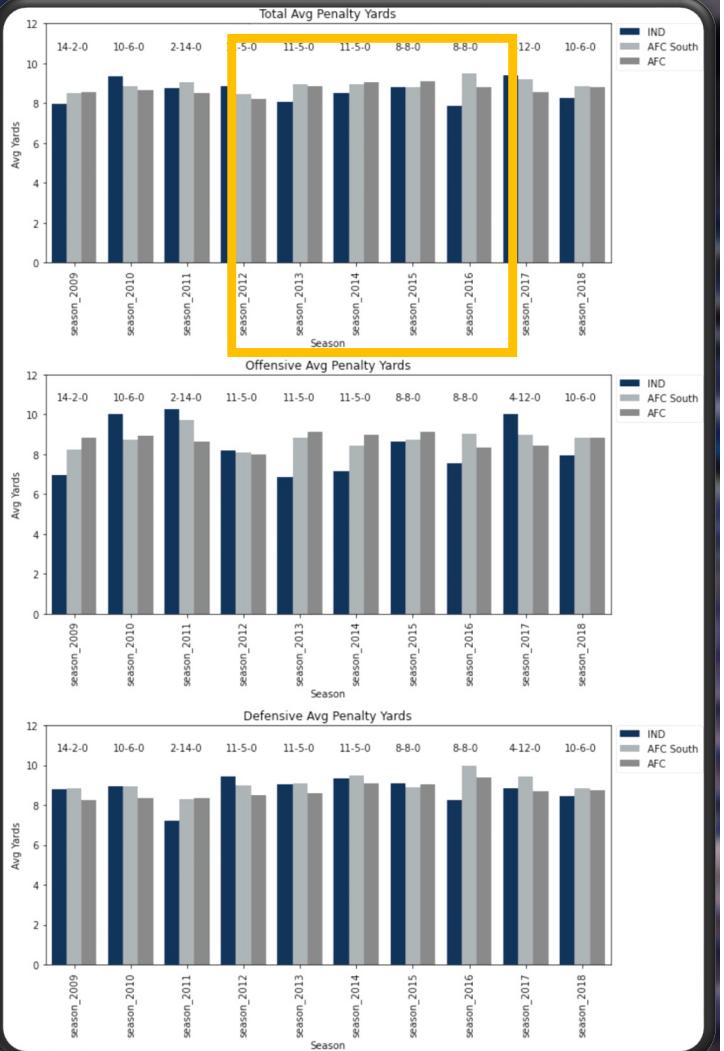
- The `avg_penalty_yards()` function returns 3 side-by-side bar charts in the team's colors.
- It will plot the total penalty yards, offensive penalty yards, and defensive penalty yards for the team passed into the function, the division the team belongs to, and the conference the team belongs to.
- Finally, the team's record for the season (the x-axis) will be displayed above each group of bars.
 - This plot is very useful to determine how a team performed (total, on offense, and on defense) against their division and conference and how successful the team was in a given season.

Tim Hulak



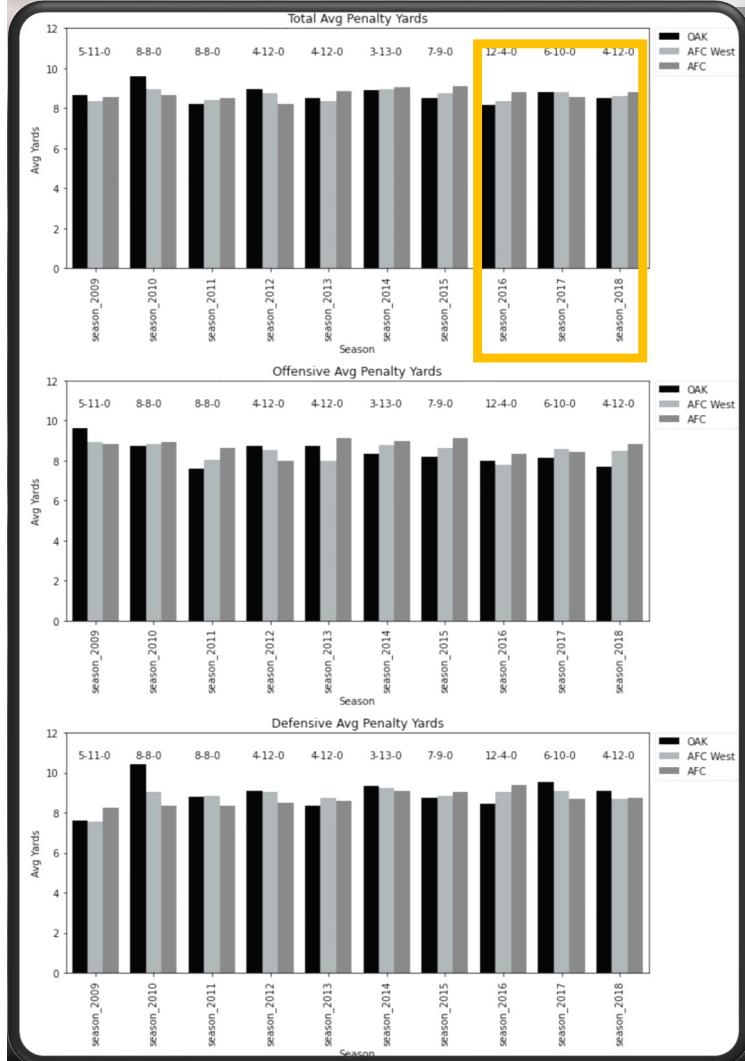
Penalties at Team Level: Indianapolis Colts

- The Colts are historically a low penalized team.
- In 2012, 2013, and 2014 the Colts had a successful record of 11-5 despite ups and downs in average penalty yards compared to their division and conference.
- In 2015 and 2016, the Colts saw an unsuccessful record of 8-8 with varying penalties compared to their division and conference



Penalties at Team Level: Oakland (Las Vegas) Raiders

- The Raiders are historically a highly penalized team.
- It can be argued that when the Raiders commit fewer penalties, they are more successful (see 2016 season with a 12-4 record)
- It can be argued that since the team is flagged so much, they lose more yards and ultimately lose games
- However, the team seems to be on par with division and conference average penalties

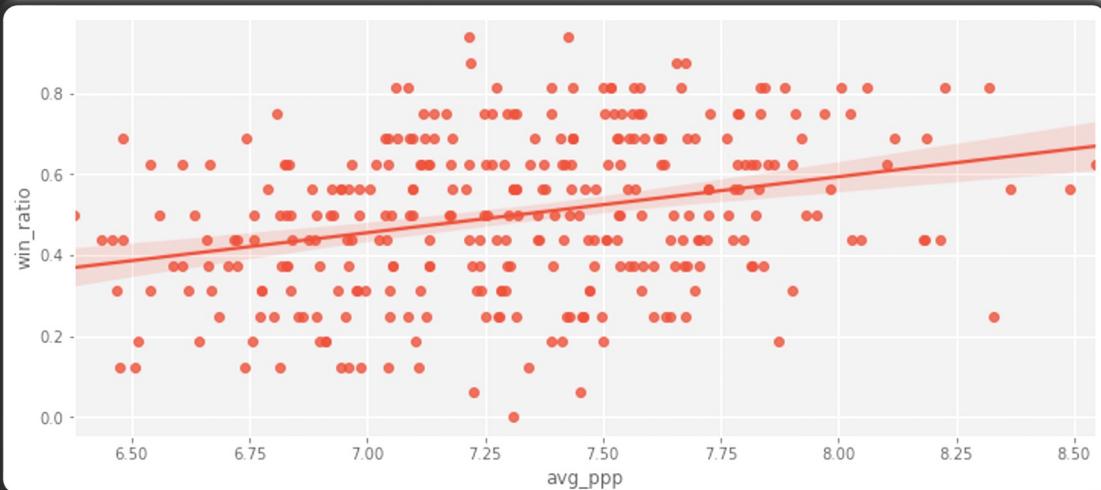


Is THERE A CORRELATION BETWEEN PLAYS PER POSSESSION AND THE SUCCESS OF A TEAM IN ANY GIVEN SEASON?

In American Football possessing the football gives a team an advantage in that it gives the possessing team the opportunity to score while denying their opponents that same opportunity. One such metric to capture this idea is “Plays Per Possession” (ppp). We postulate that this metric should correlate with team success, measured by win ratio (win_ratio), in any given season.

Jordan Hyatt

Regression Plot (avg_ppp vs win_ratio)



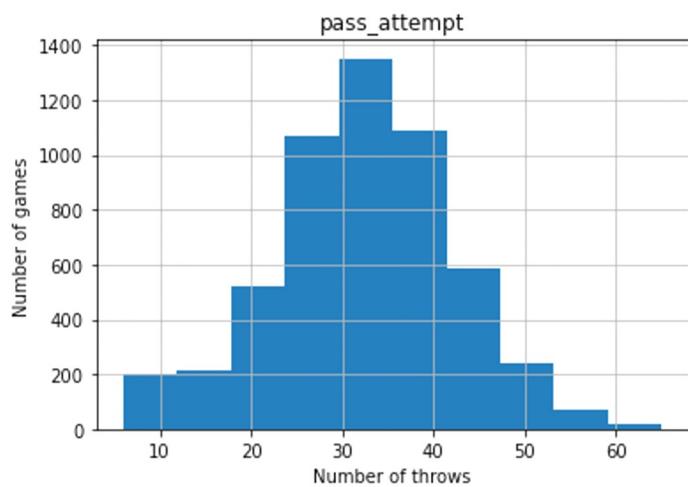
Sample of Aggregated Data

	win_ratio	season	team_abbr	avg_ppp
319	0.9375	2015	CAR	7.217172
318	0.9375	2011	GB	7.427027
317	0.8750	2016	NE	7.675676
316	0.8750	2010	NE	7.654971
315	0.8750	2009	IND	7.217877
...
4	0.1250	2010	CAR	6.507538
3	0.1250	2009	DET	6.984848
2	0.0625	2016	CLE	7.451807
1	0.0625	2009	LA	7.224044
0	0.0000	2017	CLE	7.310160

Ordinary Least Squares (OLS) Model

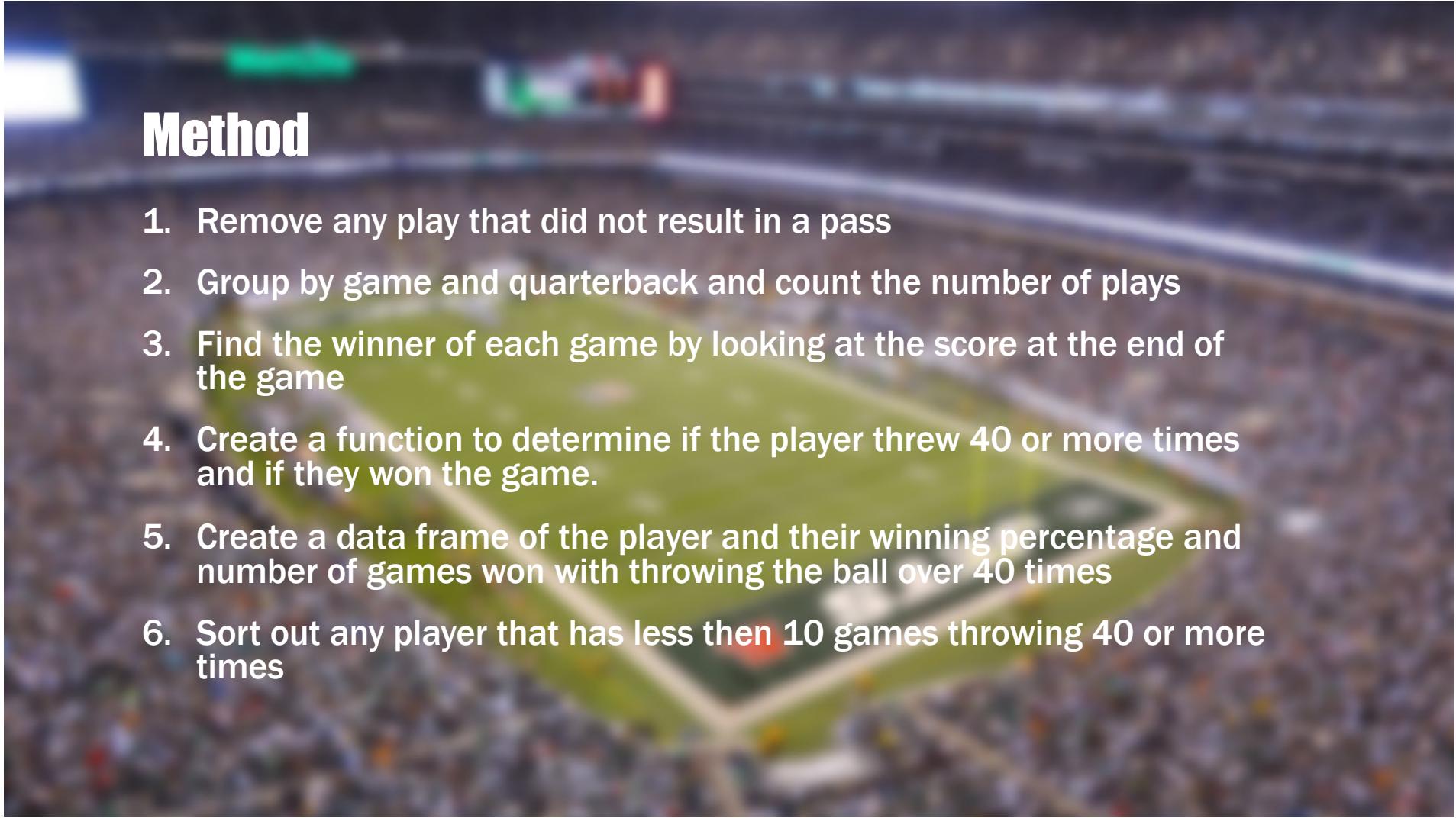
Dep. Variable:	win_ratio	R-squared (uncentered):	0.880
Model:	OLS	Adj. R-squared (uncentered):	0.879
Method:	Least Squares	F-statistic:	2330.
Date:	Thu, 02 Sep 2021	Prob (F-statistic):	1.10e-148
Time:	20:26:07	Log-Likelihood:	85.182
No. Observations:	320	AIC:	-168.4
Df Residuals:	319	BIC:	-164.6
Df Model:	1		
Covariance Type:	nonrobust		
	coef	std err	t P> t [0.025 0.975]
avg_ppp	0.0684	0.001	48.274 0.000 0.066 0.071
Omnibus:	13.461	Durbin-Watson:	0.052
Prob(Omnibus):	0.001	Jarque-Bera (JB):	6.625
Skew:	-0.112	Prob(JB):	0.0364
Kurtosis:	2.331	Cond. No.	1.00

WHAT IS THE WINNING PERCENTAGE OF QUARTERBACKS WHO THROW THE BALL (PASS) 40 OR MORE TIMES PER GAME?



- There is a debate among some that if the quarterback were to throw 40 or more times a game, it could lead to more wins
- We have a large amount that is close to throwing 40 times, so we will look at the winning percentage of those who have thrown over 40 and see if we could recommend moving the average up a bit so more quarterbacks are having 40 plus pass attempts a game.

Cal Wardell



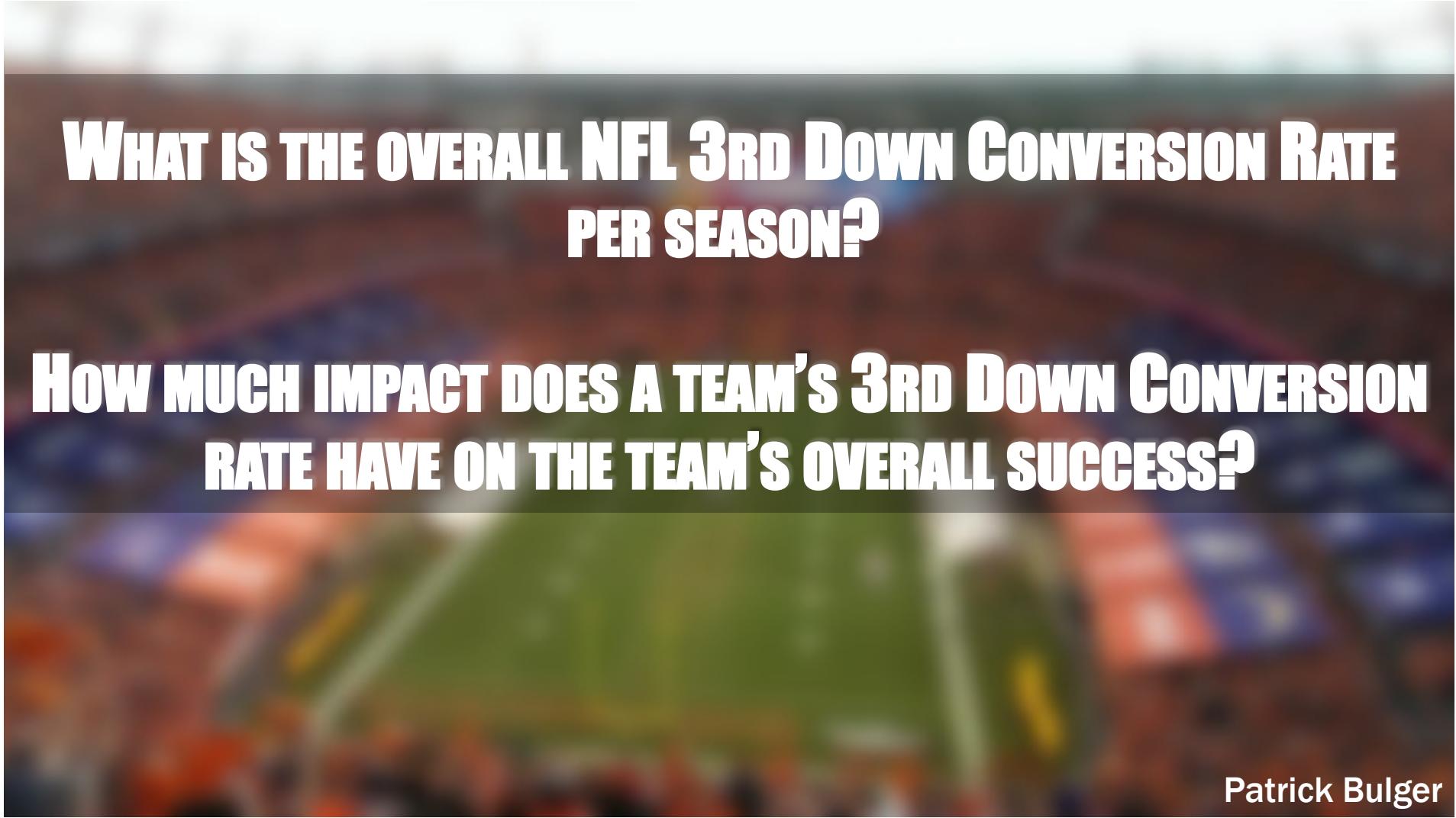
Method

1. Remove any play that did not result in a pass
2. Group by game and quarterback and count the number of plays
3. Find the winner of each game by looking at the score at the end of the game
4. Create a function to determine if the player threw 40 or more times and if they won the game.
5. Create a data frame of the player and their winning percentage and number of games won with throwing the ball over 40 times
6. Sort out any player that has less than 10 games throwing 40 or more times

		WinPerc	WinCount	NumofGames
passer_player_name	win_with_40			
T.Brady	True	0.680851	32	47.0
P.Manning	True	0.675000	27	40.0
M.Schaub	True	0.500000	7	14.0
A.Rodgers	True	0.452381	19	42.0
D.Brees	True	0.428571	30	70.0
A.Luck	True	0.400000	14	35.0
P.Rivers	True	0.365854	15	41.0
B.Roethlisberger	True	0.363636	16	44.0
A.Dalton	True	0.346154	9	26.0
J.Flacco	True	0.340909	15	44.0
M.Ryan	True	0.339623	18	53.0
M.Hasselbeck	True	0.333333	5	15.0
R.Fitzpatrick	True	0.318182	7	22.0
M.Sanchez	True	0.307692	4	13.0
E.Manning	True	0.304348	14	46.0
M.Stafford	True	0.288136	17	59.0
K.Cousins	True	0.280000	7	25.0
J.McCown	True	0.272727	3	11.0
A.Smith	True	0.263158	5	19.0
B.Hoyer	True	0.250000	3	12.0
C.Newton	True	0.230769	3	13.0
C.Wentz	True	0.230769	3	13.0
J.Freeman	True	0.230769	3	13.0
C.Palmer	True	0.222222	8	36.0
D.Carr	True	0.210526	4	19.0
J.Cutler	True	0.210526	4	19.0
C.Keenum	True	0.181818	2	11.0
T.Romo	True	0.173913	4	23.0
C.Henne	True	0.166667	3	18.0
K.Orton	True	0.153846	2	13.0
B.Bortles	True	0.117647	2	17.0
R.Tannehill	True	0.117647	2	17.0
S.Bradford	True	0.076923	2	26.0

Results

- Only two quarterbacks have a winning percentage over 50 percent.
- We have an average winning percentage of ~35 %



**WHAT IS THE OVERALL NFL 3RD DOWN CONVERSION RATE
PER SEASON?**

**HOW MUCH IMPACT DOES A TEAM'S 3RD DOWN CONVERSION
RATE HAVE ON THE TEAM'S OVERALL SUCCESS?**

Patrick Bulger

What is the total NFL 3rd Down Conversion Rate?

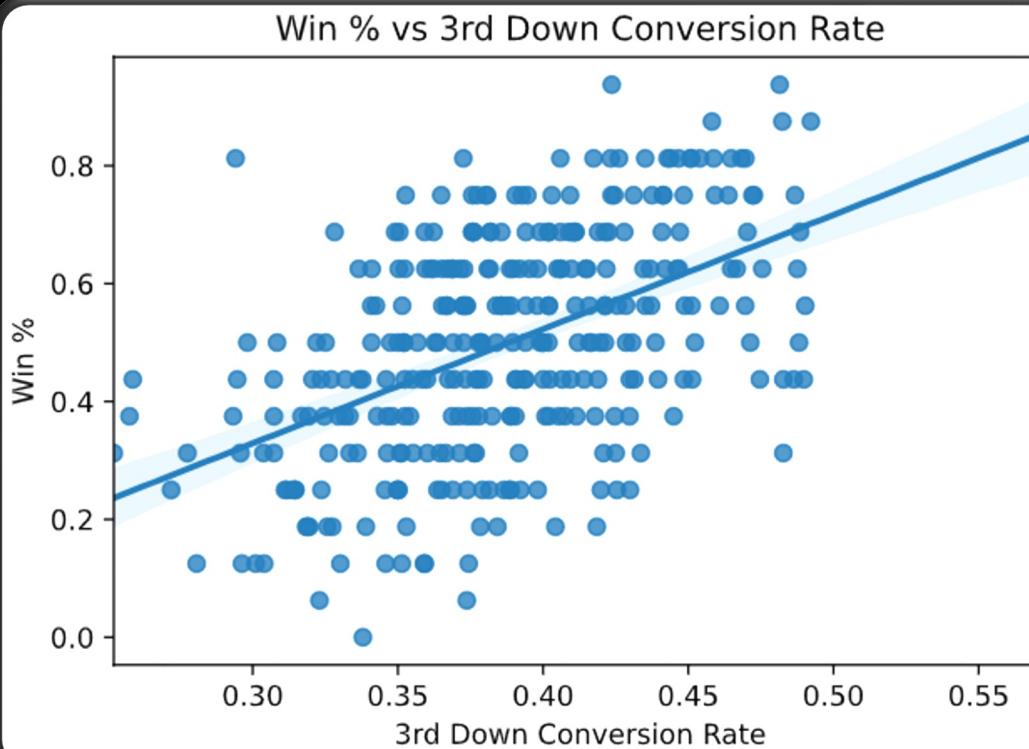
- Statisticians often speak to a team's 3rd down conversion rate as a measure of success
- To understand what a good conversion rate is, we get a benchmark rate of the season rates
- This benchmark will allow us to compare individual teams
- As we can see here, the benchmark rates per season all fall at ~39%

Season	NFL conversion rate
season_2009	0.385901
season_2010	0.382593
season_2011	0.380280
season_2012	0.380910
season_2013	0.382027
season_2014	0.399297
season_2015	0.389481
season_2016	0.397230
season_2017	0.383526
season_2018	0.392792

Win % vs 3rd Down Conversion Rate

Correlation Coefficient: 0.5056

- Calculated the winning % of each team
- Calculated the 3rd Down Conversion rate for each team
- Created a scatter plot demonstrating how they relate to each other
- Linear regression line clearly shows that as the conversion rate increases, so does the chance for winning the game
- Interesting to see some of the outliers, such as the team that had won over 80% of their games with a below benchmark conversion rate

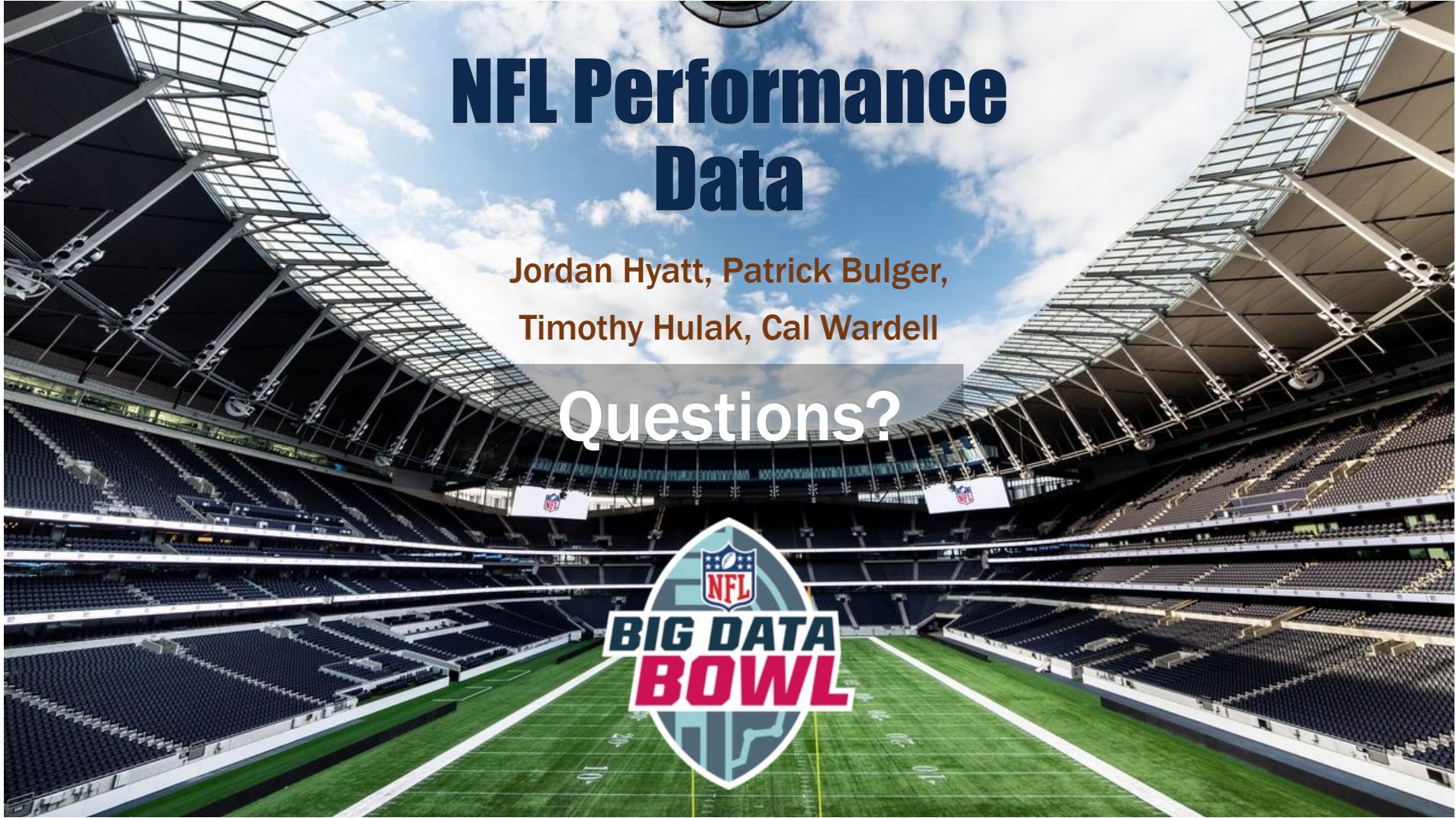


Conclusions

- **WHAT ARE THE EFFECTS OF PENALTIES ON TEAM PERFORMANCE?**
 - AT A HIGH LEVEL, TEAMS CAN FOCUS ON REDUCING THE TOP 5 VIOLATIONS IN ORDER TO BE MORE SUCCESSFUL.
 - TEAMS CAN FOCUS ON CLEANER OFFENSIVE PLAY IN ORDER TO SEE MORE SUCCESS.
 - WHILE PENALTIES CAN HAVE A DRASIC EFFECT ON A TEAM DURING A GIVEN GAME AND SITUATION, THEY MAY NOT HAVE A MACRO EFFECT ON TEAM SUCCESS WHEN TAKEN AS THE SINGLE, HIGH-LEVEL FACTOR.
- **IS THERE A CORRELATION BETWEEN PLAYS PER POSSESSION AND THE SUCCESS OF A TEAM IN ANY GIVEN SEASON?**
 - YES, ACCORDING TO OUR ANALYSIS, TEAMS THAT HAVE MORE PLAYS PER POSSESSION HAVE A HIGHER RATE OF WINNING. THERE IS A POSITIVE CORRELATION BETWEEN THESE TWO METRICS
 - THIS CORRELATION IS FURTHER BACKED UP BY 3RD DOWN CONVERSION RATE ANALYSIS WHICH HAS NON-ZERO COVARIANCE WITH PPP METRIC
 - COACHING DECISIONS AND GAME PLANS SHOULD BE GEARED TOWARD MAINTAINING POSSESSION OF THE FOOTBALL

Conclusions

- **WHAT IS THE WINNING PERCENTAGE OF QUARTERBACKS WHO THROWS THE BALL (PASS) 40 OR MORE TIMES PER GAME?**
 - THE AVERAGE WINNING PERCENTAGE OF QUARTERBACKS WITH OVER 40 PASS ATTEMPTS IS ~35%
 - **WITH ONLY 2 QUARTERBACKS WITH A WINNING PERCENTAGE(>50%) RECORD WE DO NOT RECOMMEND THAT TEAMS TRY TO INCREASE THE NUMBER OF THROWS THEIR QUARTERBACKS MAKE.**
- **HOW MUCH IMPACT DOES A TEAM'S 3RD DOWN CONVERSION RATE HAVE ON THE TEAM'S OVERALL SUCCESS?**
 - THE NFL'S BENCHMARK CONVERSION RATE IS ~39% YEAR OVER YEAR
 - THE CORRELATION BETWEEN WINNING PERCENTAGE AND CONVERSION RATE IS **0.50562**, A STRONG, POSITIVE CORRELATION
 - When plotted, we can visually see that having a good 3rd Down Conversion rate does have a positive impact on winning



NFL Performance Data

Jordan Hyatt, Patrick Bulger,
Timothy Hulak, Cal Wardell

Questions?

