Passing EPA in Outdoor NFL Games

Max S and Josh

2023-10-19

Passing EPA (Expected Points Added) in Outdoor NFL Games in the 2021 and 2022 seasons

The data was compiled using NFL play-by-play and game level data from NFLFastR (https://github.com/nflverse/nflfastR (https://github.com/nflverse/nflfastR)), cleaned and engineered using Python. The goal of this project is to predict passing EPA or passing play rate (We'll explain EPA further below) using pre game factors. The application of this model could be useful for NFL teams for game-planning. It could also have uses in sports betting and fantasy sports.

Column names and sample data below:

```
##           game_id posteam rush_rate precipitation week pass_epa_game season
## 1 2021_06_ARI_CLE     ARI 0.5151515      noprecip    6     0.4555283   2021
## 2  2021_09_ARI_SF     ARI 0.5671642 lightrainsnow    9     0.5638735   2021
## 3 2021_11_ARI_SEA     ARI 0.4025974      noprecip   11     0.3429203   2021
## 4 2021_13_ARI_CHI     ARI 0.6730769      noprecip   13     0.3945339   2021
## 5 2022_15_ARI_DEN     ARI 0.3333333      noprecip   15    -0.4375405   2022
## 6 2021_05_NYJ_ATL     ATL 0.3783784      noprecip    5     0.3404823   2021
##   weekday   surface     roof div_game spread_line total_line wind temp
## 1  Sunday    grass  outdoors        0         3.0       48.0   23   58
## 2  Sunday     grass outdoors        1         5.5       44.5    8   57
## 3  Sunday fieldturf outdoors        1         2.5       46.0    0   46
## 4  Sunday    grass  outdoors        0        -7.5       43.5    8   34
## 5  Sunday     grass outdoors        0         1.5       37.5    4   37
## 6  Sunday     grass outdoors        0        -3.0       45.5    8   63
##   epa_rush_last3 pass_rate epa_pass_last3 pass_rate_last3
## 1     0.08732964 0.4848485    0.248012093       0.5298831
## 2     0.04256709 0.4328358    0.247782096       0.5329414
## 3    -0.14714319 0.5974026   -0.004919549       0.5639470
## 4    -0.20308013 0.3269231    0.103310921       0.5509600
## 5    -0.09100651 0.6666667   -0.157502203       0.6039604
## 6    -0.26880706 0.6216216    0.078103331       0.6555503
```

```
##  [1] "game_id"         "posteam"         "rush_rate"       "precipitation"
##  [5] "week"            "pass_epa_game"   "season"          "weekday"
##  [9] "surface"         "roof"            "div_game"        "spread_line"
## [13] "total_line"      "wind"            "temp"            "epa_rush_last3"
## [17] "pass_rate"       "epa_pass_last3"  "pass_rate_last3"
```

Column Descriptions

Each row represents an individual offensive team in and individual game from the 2021 and 2022 seasons.

game_id - The identifier for a game.

posteam - Team on offense. precipitation - Categorical variable. Either no rain or snow "noprecip", light rain or snow "lightrainsnow", or moderate to heavy rain or snow "rainsnow".

Game variables (possible Y values)

rush_rate - Rate of rushing plays per offensive play for the game.

pass_rate - Rate of passing plays per offensive play for the game.

pass_epa_game - Average expected points added per pass play for the posteam over the full game. From NFLFastR "Expected Points (EP): using scoring event probabilities, the estimated expected points with respect to the possessing team before the play; simply put, the value of a particular situation to the offensive team, measured by points. Expected Points Added (EPA): expected points after the play minus expected points before." This statistic is the best way to measure success for an NFL offense.

Pre-game variables (possible X values)

weekday - Day of the week.

surface - NFL Stadiums have either natural grass or turf.

div_game - Categorical variable 1 or 0. 1 represents a game played by two teams in the same division. Division games usually have bigger playoff implications and longer standing rivalries.

spread_line - The consensus sports book "spread." A negative spread means the team is favored, a positive spread represents an underdog.

total_line - The consensus sports book "total." The total represents the over under for the total points predicted to be scored by sports bettors in an NFL game.

wind - wind speed in miles per hour at the start of the game.

temp - temperature (Fahrenheit) at the start of the game.

epa_rush_last3 - average EPA on rushing plays for the team in their last 3 games.

epa_pass_last3 - average EPA on passing plays for the team in their last 3 games.

pass_rate_last3 - passing play rate for the team in their last 3 games.

Summary of the data:

Data summary

Name	data
Number of rows	448
Number of columns	19
_____	
Column type frequency:	
character	6
numeric	13
_____	
Group variables	None

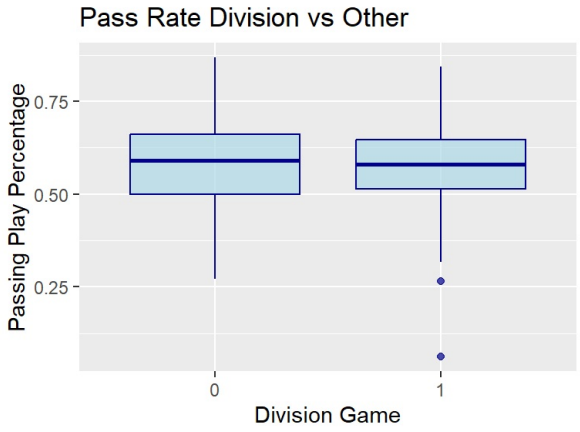
Variable type: character

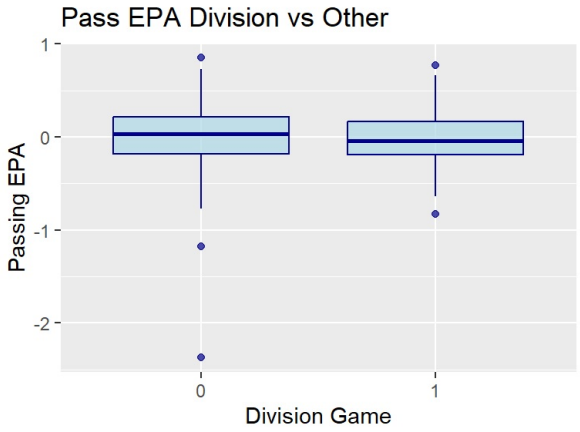
skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
game_id	0	1	13	15	0	224	0
posteam	0	1	2	3	0	32	0
precipitation	0	1	8	13	0	3	0
weekday	0	1	6	8	0	5	0
surface	0	1	5	9	0	4	0
roof	0	1	8	8	0	1	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
rush_rate	0	1	0.42	0.11	0.13	0.34	0.41	0.50	0.94	
week	0	1	11.52	3.84	4.00	9.00	12.00	15.00	17.00	
pass_epa_game	0	1	0.01	0.32	-2.37	-0.19	0.01	0.21	0.85	
season	0	1	2021.38	0.48	2021.00	2021.00	2021.00	2022.00	2022.00	
div_game	0	1	0.32	0.47	0.00	0.00	0.00	1.00	1.00	
spread_line	0	1	0.00	6.89	-19.00	-4.00	0.00	4.00	19.00	
total_line	0	1	44.53	4.21	32.00	42.00	44.25	47.00	57.50	
wind	0	1	9.11	5.62	0.00	6.00	9.00	12.00	27.00	
temp	0	1	52.04	16.78	6.00	41.00	53.00	62.25	88.00	
epa_rush_last3	0	1	-0.02	0.12	-0.48	-0.10	-0.02	0.06	0.41	
pass_rate	0	1	0.58	0.11	0.06	0.50	0.59	0.66	0.87	
epa_pass_last3	0	1	0.02	0.19	-0.47	-0.11	0.02	0.16	0.55	
pass_rate_last3	0	1	0.58	0.08	0.33	0.53	0.58	0.63	0.76	

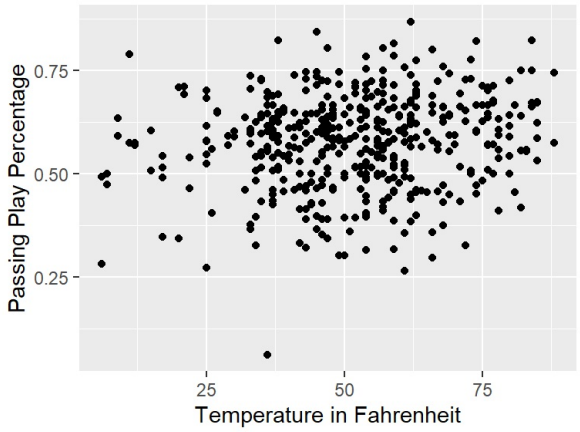
Teams seem to pass less in division games and also have less success passing.

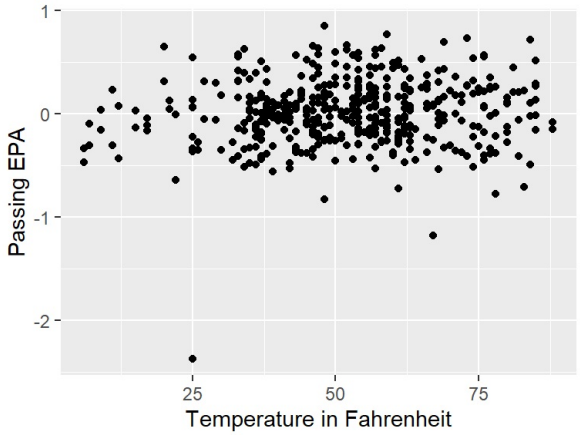




Teams seem to be more conservative in divisional games.

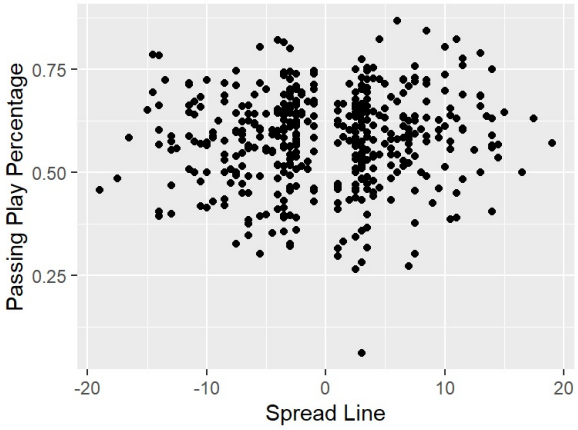
Temperature also seems to have a relationship with passing rate but not as much with passing EPA.

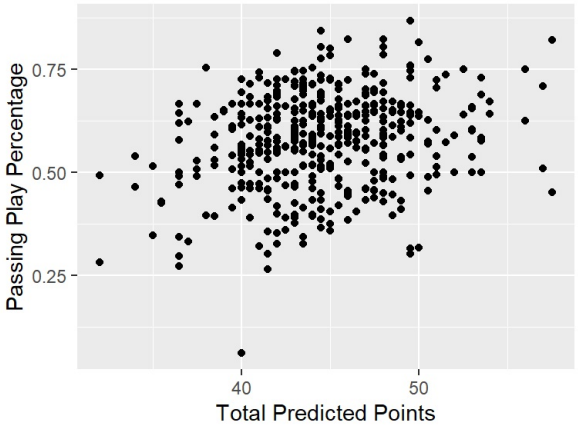




Like wind, conventional wisdom is that teams pass less and with less success in colder temps, but the data seems to have a lot of variance. Might make sense to change into a categorical variable.

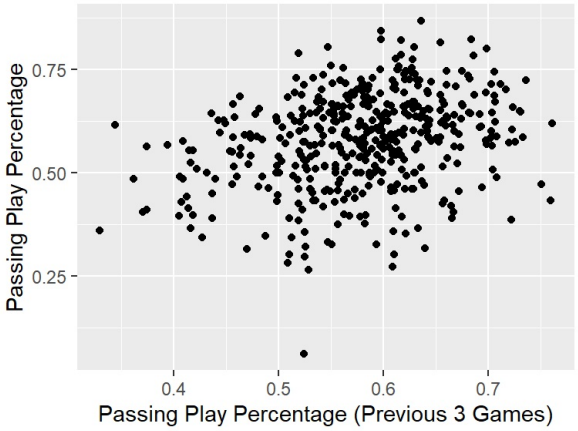
Spread line does not seem to have a relationship with passing rate but total line seems to have a strong positive relationship

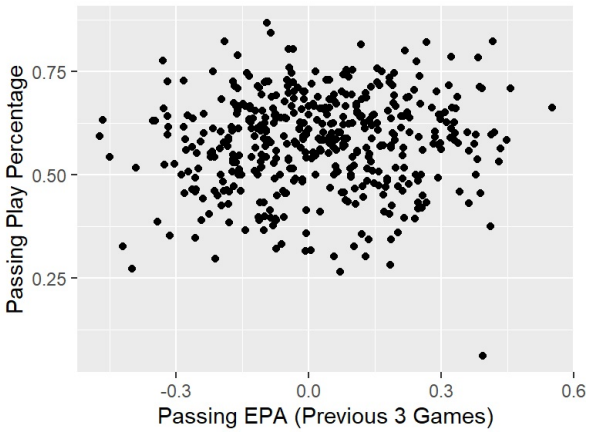




This could lead us to the interesting conclusion that while passing the football may lead to scoring more points, it does not necessarily correlate to winning more games. The typical rationale would be that scoring more points leads to winning more games, but that is perhaps not the case.

A teams passing rate in their last 3 games seems to have a strong relationship with their passing rate in the following game, but passing success (EPA) in their last 3 games does not.





Conclusions could be more about team psychology here, regardless of passing performance over the last 3 games they will continue to pass. Perhaps could be applicable to defensive strategy, defenses don't need to worry about productivity, only patterns of actions.