Temple Football Play-by-Play Analysis

Sam Ouellette, Gavin Wood, Ramdhan Russell, Jeremy Mathew

Competition Prompt

- Using the season-long play-by-play data, select either OFFENSE OR DEFENSE as your focus. Your goal is to determine the factors that most strongly influence successful outcomes on that side of the ball. You must define what “success” means (e.g., scoring drives, preventing first downs, preventing red zone entry, etc.) and provide a clear justification of your definition using evidence from the dataset.
- You are free to decide which variables to include and why, but you must defend your choices using football logic and data support. After determining the drivers of success, develop a recommendation plan that shows how the team should adjust sequencing and decision tendencies to improve success on the side you selected. This means recommending how the team should change when and how often they choose certain play types to maximize success. For example, should they run more on 2nd and short, pass more on early downs, or be more aggressive in midfield situations? Your recommendations should be grounded in what your analysis shows leads to better outcomes.
- Your final submission must include descriptive analytics, visual evidence, and a predictive model (any model type is acceptable, if it is justified) to support your recommendation.

Overview

- We built two main models:
  1. Logistic regression model that analyzes which factors are most important in predicting a successful play (positive EPA)
  2. Formation performance model that shows the average number of yards gained and the average EPA for each type of offensive formation

What is EP and EPA?

- What is EP and EPA?
  - EP = Expected points
    - Average points a team is expected to score on the current drive given down, distance, and field position
    - Example: If teams with 1st and 10 at their own 20-yard line historically average 1 expected point, then teams typically finish the drive scoring 1 point
      - (they don’t score 1 point. It's an average based on data from thousands of drives)

What is EPA?

- EPA= Expected points added
  - Change in expected points from start of a play to the end
  - You gain 5 yards on first and 10 to have 2nd and 5 on your own 25. This new down and distance has an EP of 1.5. The 5-yard play that you ran has an EPA of 0.5

This project defines success as any play with an EPA>0 as these plays will always increase your expected points scored on the drive

Logistic Regression EPA Model

- Ran logistic regression model to see which play factors most strongly influence EPA
- Response is binary EPA. Response=1 when EPA>0 (success) and response=0 when EPA<=0
- Factors chosen for model:
  - Down
  - Distance
  - Run/pass play
  - Yard to endzone
  - Left/right hash
  - Score differential
  - Quarter number
  - Time remaining in game (in seconds)
  - Side of field
  - Previous play EPA (a bigger or smaller play can affect momentum for nect play)
  - Run/pass on previous play
  - Play # of each drive

Model equation:

```
Call:
glm(formula = success ~ DOWN + DISTANCE + RUN_PASS + yards_to_endzone +
    HASH + SCORE_DIFF + QUARTER + time_remaining + FIELD_SIDE +
    prev_epa + prev_run_pass + DRIVE_PLAY, family = binomial(),
    data = df)
```

Model Output

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.8612113  1.0285327  -1.810   0.0704 .
DOWN               0.0929225  0.1263565   0.735   0.4621
DISTANCE           0.0339334  0.0236665   1.434   0.1516
RUN_PASSR         -0.3174559  0.1572310  -2.019   0.0435 *
RUN_PASSX         -2.5024862  1.0495707  -2.384   0.0171 *
yards_to_endzone  -0.0038242  0.0066487  -0.575   0.5652
HASHL             -0.2813546  0.2101997  -1.339   0.1807
HASHR              0.0759358  0.2094311   0.363   0.7169
SCORE_DIFF        -0.0072148  0.0058211  -1.239   0.2152
QUARTER2           0.3418365  0.3329272   1.027   0.3045
QUARTER3           0.9649185  0.5481818   1.760   0.0784 .
QUARTER4           1.4727250  0.8038620   1.832   0.0669 .
QUARTER5           0.1156063  1.4369787   0.080   0.9359
time_remaining     0.0006029  0.0002854   2.113   0.0346 *
FIELD_SIDEOWN      0.1283086  0.2804630   0.457   0.6473
prev_epa           0.0759779  0.1022634   0.743   0.4575
prev_run_pass2     0.0849442  0.1702660   0.499   0.6179
prev_run_pass3     0.3588314  0.6297473   0.570   0.5688
prev_run_passNONE -0.2563861  0.2535777  -1.011   0.3120
DRIVE_PLAY        -0.0435615  0.0385671  -1.130   0.2587
```

Significant Factors (determined from p-values):

- Run/pass play
- Third Quarter
- Fourth Quarter
- Time Remaining in game (seconds)

Estimate column shows the log odds of success for that variable when controlling all other variables. Taking the exp of the log odds will give you the odds of success, or odds ratio

For RUN_PASSR:

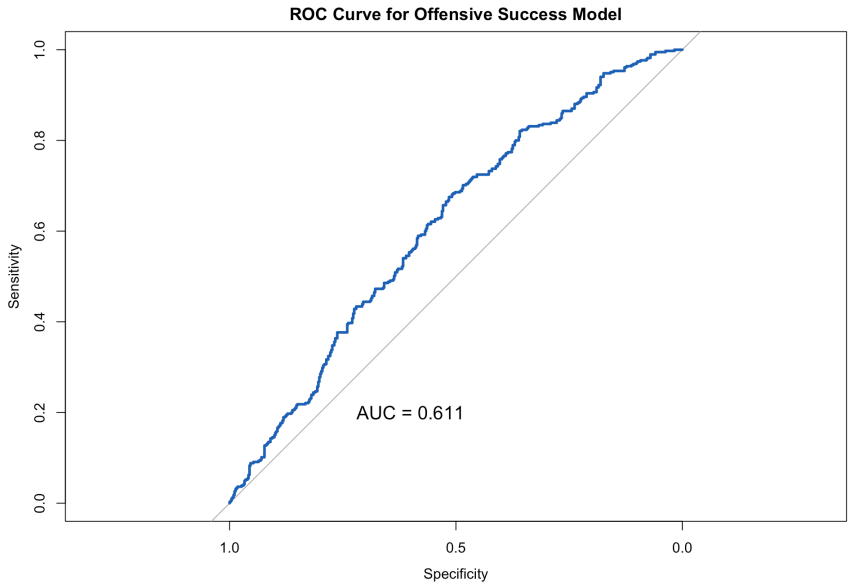
- Exp(-0.317)= 0.729
- Translation: Choosing to run rather than pass multiplies odds of success by 72.9%, or decreases the odds of success by 27%

Full Factor Interpretations

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.8612113  1.0285327  -1.810   0.0704 .
DOWN              0.0929225  0.1263565   0.735   0.4621
DISTANCE          0.0339334  0.0236665   1.434   0.1516
RUN_PASSR        -0.3174559  0.1572310  -2.019   0.0435 *
RUN_PASSX        -2.5024862  1.0495707  -2.384   0.0171 *
yards_to_endzone -0.0038242  0.0066487  -0.575   0.5652
HASHL            -0.2813546  0.2101997  -1.339   0.1807
HASHR             0.0759358  0.2094311   0.363   0.7169
SCORE_DIFF       -0.0072148  0.0058211  -1.239   0.2152
QUARTER2          0.3418365  0.3329272   1.027   0.3045
QUARTER3          0.9649185  0.5481818   1.760   0.0784 .
QUARTER4          1.4727250  0.8038620   1.832   0.0669 .
QUARTER5          0.1156063  1.4369787   0.080   0.9359
time_remaining    0.0006029  0.0002854   2.113   0.0346 *
FIELD_SIDEOWN     0.1283086  0.2804630   0.457   0.6473
prev_epa          0.0759779  0.1022634   0.743   0.4575
prev_run_pass2    0.0849442  0.1702660   0.499   0.6179
prev_run_pass3    0.3588314  0.6297473   0.570   0.5688
prev_run_passNONE -0.2563861  0.2535777  -1.011   0.3120
DRIVE_PLAY        -0.0435615  0.0385671  -1.130   0.2587
```

- RUN vs PASS
  - Odds ratio: 0.73
  - Run plays have 27 percent lower odds of posting positive EPA than pass plays
  - Takeaway: Passing is more efficient than running
- QUARTER 3
  - Odds ratio: 2.63
  - Plays in Quarter 3 are about 2.6× more likely to be successful than in Quarter 1
  - Takeaway: Offenses come out of halftime more efficient after making adjustments.
- QUARTER 4
  - Odds ratio: 4.36
  - Plays in Quarter 4 are over 4× more likely to be successful than in Quarter 1
  - Takeaway: Late-game offense as a whole becomes more effective (tempo, urgency, softer coverages).
- TIME REMAINING (Seconds left in the GAME)
  - Odds ratio: 1.0006 per second
  - Each additional second on the game clock increases success odds by about 0.06 percent.
  - Over about 16 minutes, that becomes an 82 percent total increase in odds.
  - Takeaway: Efficiency drops only in the final minutes when time is very low and play-calling becomes predictable.

**Temple operates more efficiently in the 3rd and 4th quarters overall, but efficiency still drops when the game clock gets extremely low.



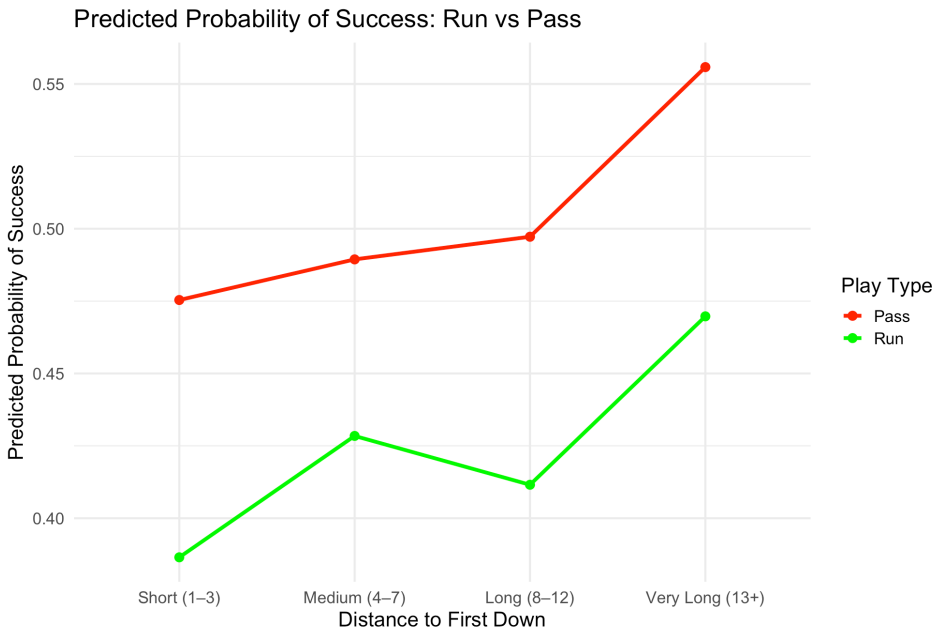
The ROC curve measures the model's ability to distinguish successful plays over unsuccessful plays.

Blue line → model's predictive performance

Gray line → random guessing

Model predicts a successful play over an unsuccessful play 61% of the time, demonstrating modest predictive power

This is normal for an EPA model, as offensive play success is noisy and depends on many things not in the data (blocking, pressure, individual matchups, QB reads, etc.)



•Pass plays have a consistently higher predicted probability of success at *every* distance bucket.

•The gap between pass and run widens as distance increases, meaning passes scale better with long-yardage situations.

•Even on short yardage, passes slightly outperform runs in terms of expected success.

- •(likely due to defenses expecting a run on down and short)

•runs become significantly less effective in down and long situations

**important to remember that success is defined as positive EPA, which is why probability increases at very long distances

- Example: a 2nd and 14 play that results in 3rd and 6 would produce a positive EPA

Coach's Recommendations

1. Pass More in Neutral situations
    - Run plays generate 27% lower odds of success
2. Focus heavily on halftime adjustments
    - Q3 shows 160% increase in success odds compared to Q1
    - Temple benefits significantly from halftime adjustments
3. Attack defenses aggressively in Q4
    - Q4 success odds are 4 times higher than Q1
    - Defenses are often fatigued and have softer coverages
    - Focus on tempo and vertical concepts (passing is more successful than running)
4. Avoid crunch-time collapse
    - Success drops sharply late in game
    - Begin late-game tempo earlier than just final couple of minutes
5. Maintain clock flexibility
    - Higher success with more time remaining
    - Staying ahead of sticks reduces need for rushed, low-efficiency plays
6. Use early down passing to unlock play action
    - Passing more successful than running
    - Stress defenses with early pass plays
    - Use play action once defenses retreats
7. Stick to passing on down & long
8. Strategically use the run, don't rely on it
    - Passing is much more successful; implement the run to keep defenses honest

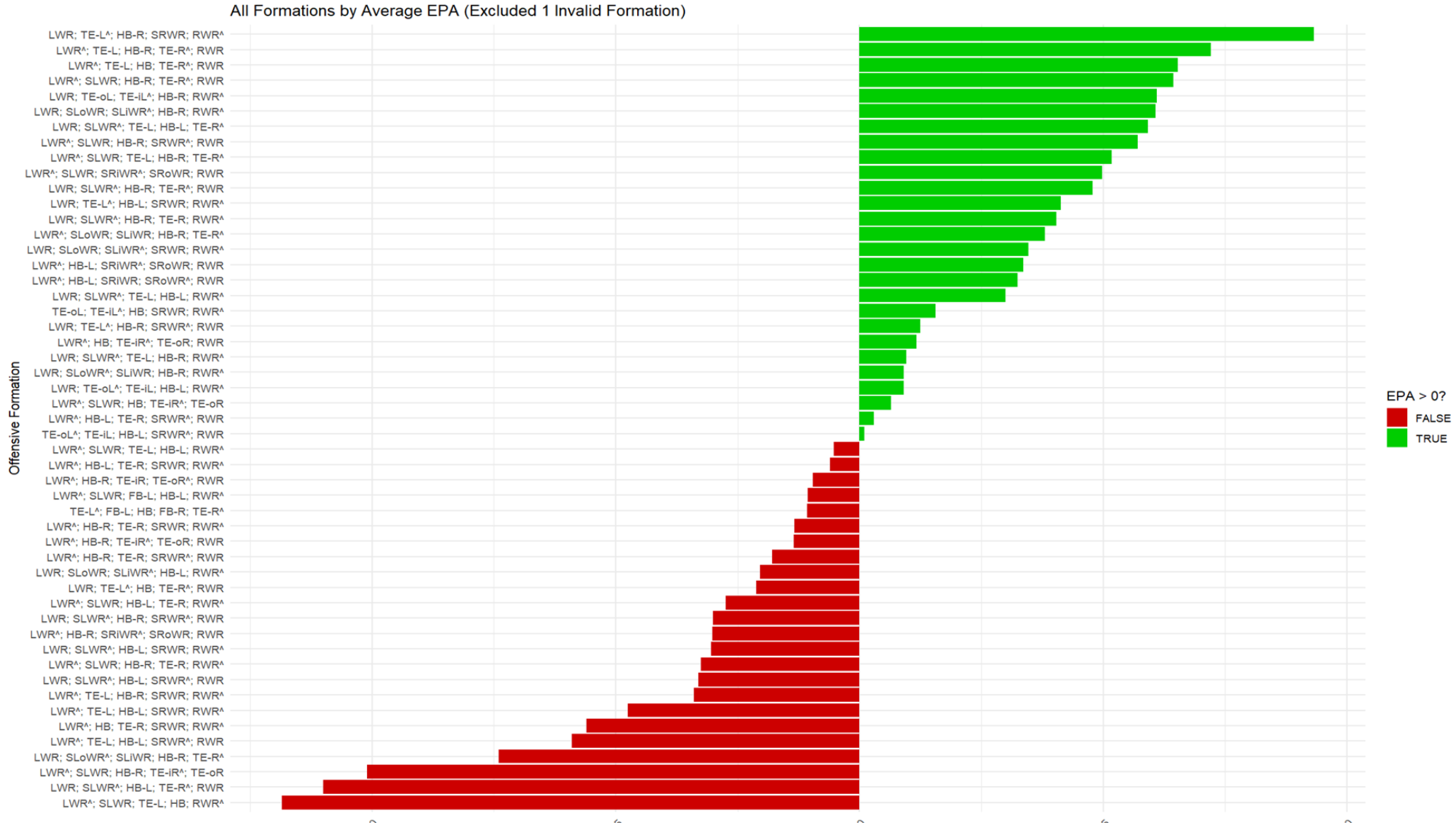
Formation Performance Model

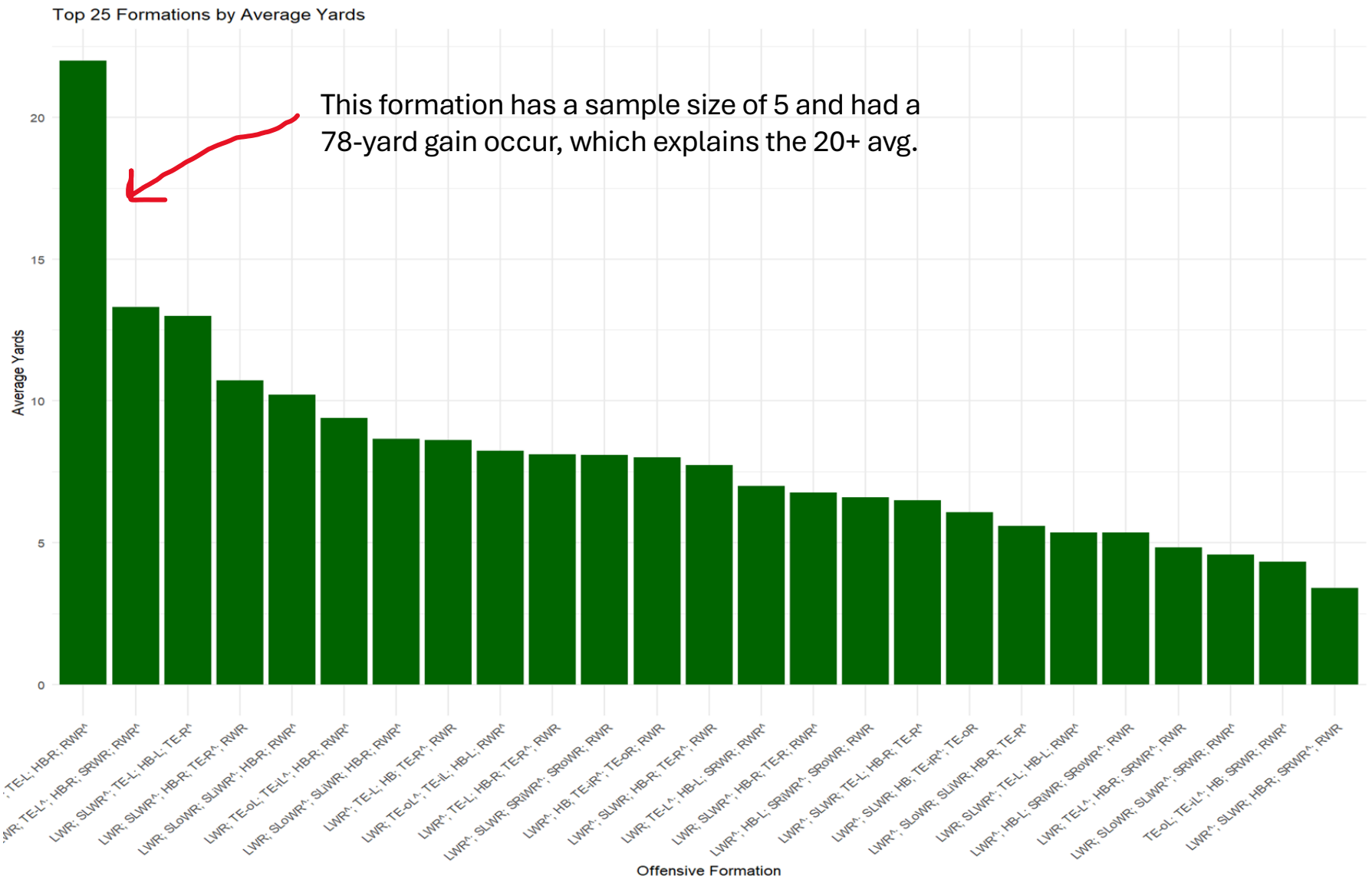
- This model aims to:
  - Analyze Temple's offensive plays by formation(exact) and formation(grouped).
    - **Formation(exact) example:** LWR; SLoWR; SLiWR^; SRWR; RWR^
    - **Formation(grouped) example:** TRIPS, DOUBLES, TRIPLE, DOLLY
  - Calculate average yards gained and expected points added (EPA)
  - Highlight the most effective formations and personnel groupings
  - Visualize performance to inform strategic play decisions

Formation(exact) Context

- For LWR; SLoWR; SLiWR^; SRWR; RWR^:
  - LWR = Left Wide Receiver(out wide)
  - SLoWR = Slot Left Outside Wide Receiver
  - SLiWR^ = Slot Left Inside Wide Receiver
  - SRWR = Slot Right Wide Receiver
  - RWR^ = Right Wide Receiever(out wide)

  - ^ = on the line of scrimmage


All Formations by Average EPA (Excluded 1 Invalid Formation)
Offensive Formation
LWR; TE-L^; HB-R; SRWR; RWR^
LWR^; TE-L; HB-R; TE-R^; RWR
LWR^; TE-L; HB; TE-R^; RWR
LWR^; SLWR; HB-R; TE-R^; RWR
LWR; TE-oL; TE-iL^; HB-R; RWR^
LWR; SLoWR; SLiWR^; HB-R; RWR^
LWR; SLWR^; TE-L; HB-L; TE-R^
LWR^; SLWR; HB-R; SRWR^; RWR
LWR^; SLWR; TE-L; HB-R; TE-R^
LWR^; SLWR; SRiWR^; SRoWR; RWR
LWR; SLWR^; HB-R; TE-R^; RWR
LWR; TE-L^; HB-L; SRWR; RWR^
LWR; SLWR^; HB-R; TE-R; RWR^
LWR^; SLoWR; SLiWR; HB-R; TE-R^
LWR; SLoWR; SLiWR^; SRWR; RWR^
LWR^; HB-L; SRiWR^; SRoWR; RWR
LWR^; HB-L; SRiWR; SRoWR^; RWR
LWR; SLWR^; TE-L; HB-L; RWR^
TE-oL; TE-iL^; HB; SRWR; RWR^
LWR; TE-L^; HB-R; SRWR^; RWR
LWR^; HB; TE-iR^; TE-oR; RWR
LWR; SLWR^; TE-L; HB-R; RWR^
LWR; SLoWR^; SLiWR; HB-R; RWR^
LWR; TE-oL^; TE-iL; HB-L; RWR^
LWR^; SLWR; HB; TE-iR^; TE-oR
LWR^; HB-L; TE-R; SRWR^; RWR
TE-oL^; TE-iL; HB-L; SRWR^; RWR
LWR^; SLWR; TE-L; HB-L; RWR^
LWR^; HB-L; TE-R; SRWR; RWR^
LWR^; HB-R; TE-iR; TE-oR^; RWR
LWR^; SLWR; FB-L; HB-L; RWR^
TE-L^; FB-L; HB; FB-R; TE-R^
LWR^; HB-R; TE-R; SRWR; RWR^
LWR^; HB-R; TE-iR^; TE-oR; RWR
LWR^; HB-R; TE-R; SRWR^; RWR
LWR; SLoWR; SLiWR^; HB-L; RWR^
LWR; TE-L^; HB; TE-R^; RWR
LWR^; SLWR; HB-L; TE-R; RWR^
LWR; SLWR^; HB-R; SRWR^; RWR
LWR^; HB-R; SRiWR^; SRoWR; RWR
LWR; SLWR^; HB-L; SRWR; RWR^
LWR^; SLWR; HB-R; TE-R; RWR^
LWR; SLWR^; HB-L; SRWR^; RWR
LWR^; TE-L; HB-R; SRWR; RWR^
LWR^; TE-L; HB-L; SRWR; RWR^
LWR^; HB; TE-R; SRWR; RWR^
LWR^; TE-L; HB-L; SRWR^; RWR
LWR; SLoWR^; SLiWR; HB-R; TE-R^
LWR^; SLWR; HB-R; TE-iR^; TE-oR
LWR; SLWR^; HB-L; TE-R^; RWR
LWR^; SLWR; TE-L; HB; RWR^
EPA > 0?
FALSE
TRUE


Top 25 Formations by Average Yards
This formation has a sample size of 5 and had a 78-yard gain occur, which explains the 20+ avg.
Average Yards
0
5
10
15
20
Offensive Formation

Most Efficient Formation

DON is the most efficient formation that Temple has in its playbook.

FORMATION(SPECIFIC)	AVG_YARDS	AVG_EPA	N_PLAYS	FORMATION(GROUPED)
LWR; TE-L^; HB-R; SRWR; RWR^	13.3	0.932	10	DON
LWR; SLWR^; HB-R; TE-R^; RWR	10.7	0.478	11	DON
LWR^; SLWR; HB-R; TE-R^; RWR	7.73	0.644	15	DON
LWR; TE-L^; HB-L; SRWR; RWR^	7	0.413	12	DON
LWR; TE-L^; HB-R; SRWR^; RWR	4.83	0.125	6	DON
LWR; SLWR^; HB-L; TE-R^; RWR	1.33	-1.1	6	DON

Other Formations to Keep on Game Script

TRIPLE and TRIPS are both formations that are able to produce both great avg_yards and avg_epa.

FORMATION(SPECIFIC)	AVG_YARDS	AVG_EPA	N_PLAYS	FORMATION(GROUPED)
LWR; SLoWR; SLiWR^; HB-L; RWR^	2.29	-0.204	7	TRIPLE
LWR; SLoWR^; SLiWR; HB-R; RWR^	8.67	0.0911	9	TRIPLE
LWR; SLoWR; SLiWR^; HB-R; RWR^	10.2	0.607	18	TRIPLE
LWR^; HB-R; SRiWR^; SRoWR; RWR	2.36	-0.302	11	TRIPS
LWR^; HB-L; SRiWR; SRoWR^; RWR	5.36	0.324	14	TRIPS
LWR^; HB-L; SRiWR^; SRoWR; RWR	6.61	0.336	28	TRIPS

Formations to Phase Out

- These are some formations that don't have good performance when it comes to avg_yards and avg_epa per play.
  - Formations:
    - DUO
    - LWR^; HB-R; TE-R; SRWR; RWR^

Formation(specific)	avg_yards	avg_epa	n_plays	Formation
LWR^; TE-L; HB-L; SRWR^; RWR	-2.14	-0.59	7	DUO
LWR^; TE-L; HB-L; SRWR; RWR^	5	-0.476	21	DUO
LWR^; TE-L; HB-R; SRWR; RWR^	1.44	-0.34	9	DUO

Formation(specific)	avg_yards	avg_epa	n_plays
LWR^; HB-R; TE-R; SRWR; RWR^	2.72	-0.134	46

The main plays called in this formation above is inside/outside zone runs or passes with either screen-flat/slant flat concepts.

Final Recommendation

Formations to continue/increase frequency:

-DON(most efficient)

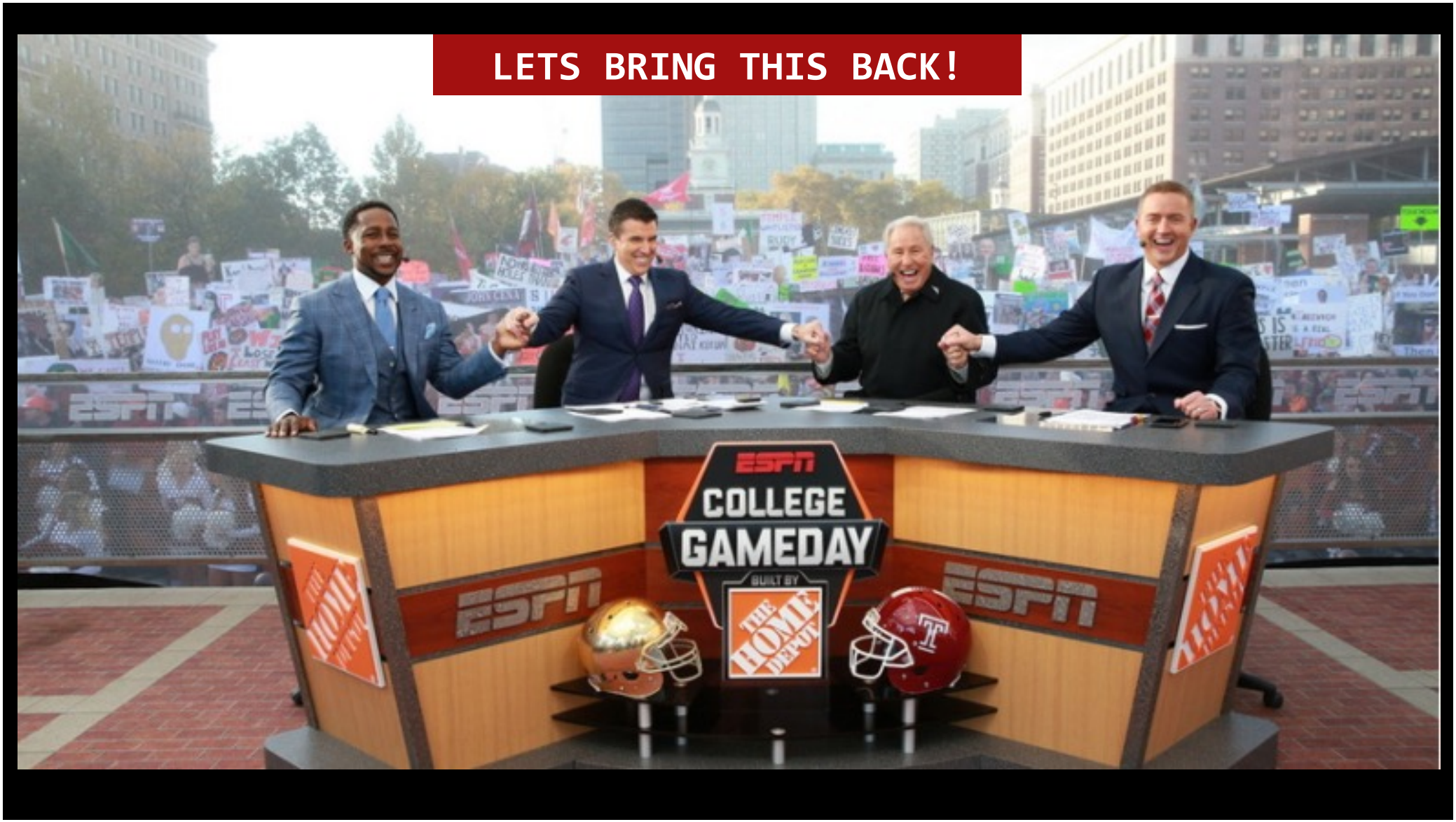
-TRIPLE

-TRIPS

Formations to phase out/run less often:

-DUO

-LWR^; HB-R; TE-R;SRWR; RWR^(personnel specific)


LETS BRING THIS BACK!
ESPN
COLLEGE
GAMEDAY
BUILT BY
THE HOME DEPOT
ESPN
ESPN