

GO Ball Analysis - 2026 NFL Big Data Bowl

Analysis of how defenses defend deep vertical "go" routes and how offenses adjust route stems, spacing, and timing to attack different coverage structures.

Project Overview

- **Data:** NFL Next Gen Stats tracking data for passing plays with vertical concepts (2023 season)
 - **Methods:** Route labeling, separation and leverage metrics, coverage identification heuristics, outcome modeling
 - **Goal:** Quantify how defensive structures influence the success of deep routes and identify tendencies teams can exploit
-

Key Questions

1. **What makes a GO route successful?**
 2. **Are there distinct types of GO routes?**
 3. **Can we predict completion probability?**
 4. **How do defenders recover on deep routes?**
 5. **Where on the field are GO routes most effective?**
-

Dataset

- **1,331 GO routes** analyzed (completions and incompletions only, no penalties)
- **577 completions** (43.4%)
- **754 incompletions** (56.6%)
- All 18 weeks of 2023 NFL season

Metric Definitions

Metric	Definition
separation_at_throw	Distance (yards) between receiver and nearest defender at moment of throw
separation_at_arrival	Distance (yards) between receiver and nearest defender when ball arrives
closing_speed	Rate at which defender closed the gap (yards/second)
air_yards	Depth of throw beyond line of scrimmage
lateral_deviation	Y-axis movement during route (indicates route adjustment)

Note: "Nearest defender" is defined as whoever is physically closest to the receiver at pass arrival, then tracked back to the throw moment.

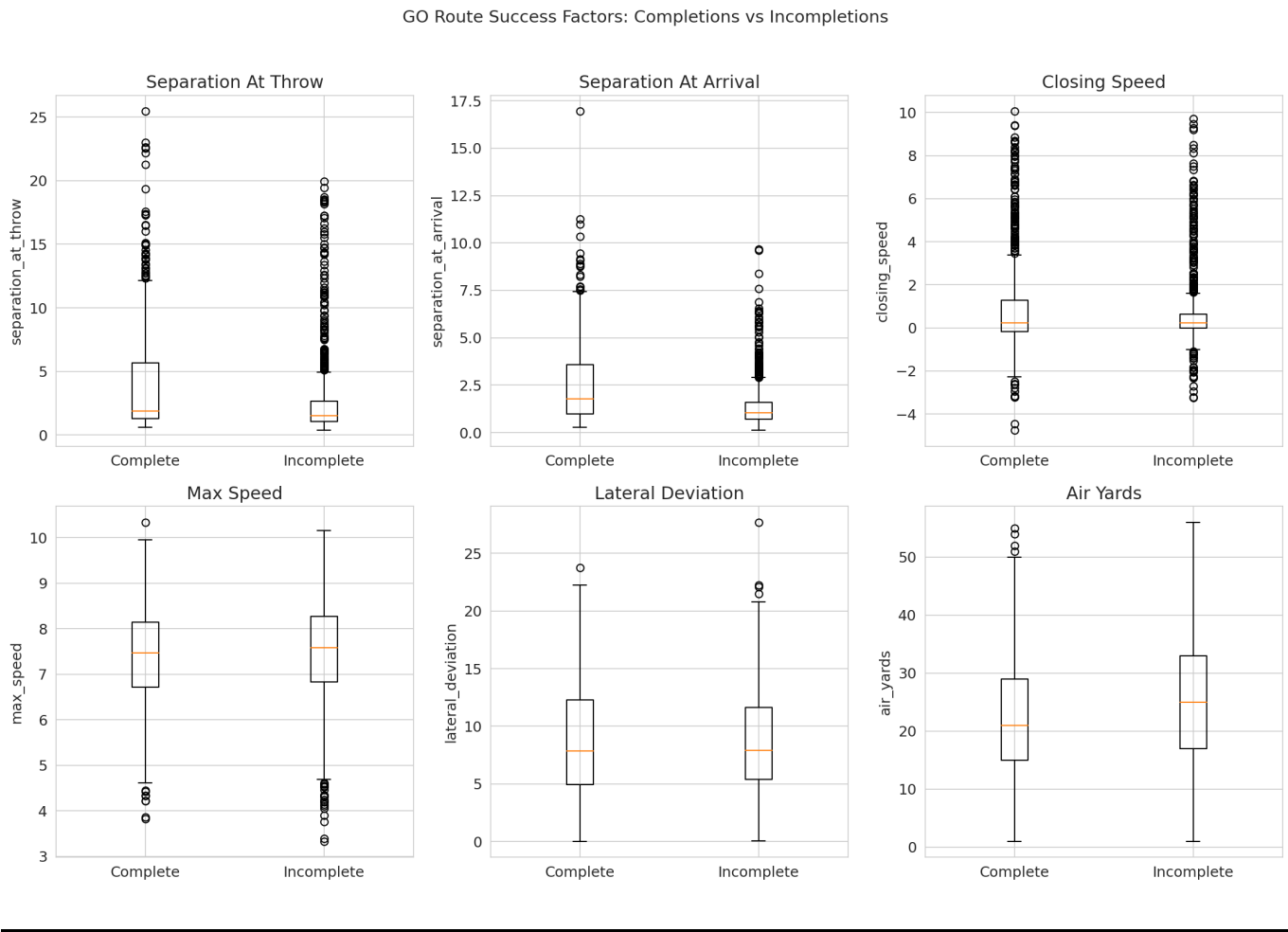
Results

Question 1: What Makes a GO Route Successful?

Separation is the strongest predictor of success.

Metric	Completions	Incompletions	Difference	p-value
Separation at Throw	4.07 yds	2.78 yds	+1.28 yds	<0.001***
Separation at Arrival	2.51 yds	1.43 yds	+1.09 yds	<0.001***
Air Yards	22.6 yds	25.4 yds	-2.78 yds	<0.001***
Closing Speed	1.03 yd/s	0.70 yd/s	+0.33 yd/s	0.003**
Max Speed	7.41 yd/s	7.46 yd/s	-0.05 yd/s	0.447

Key Insight: Receiver speed doesn't significantly differ between outcomes - it's about creating and maintaining separation.

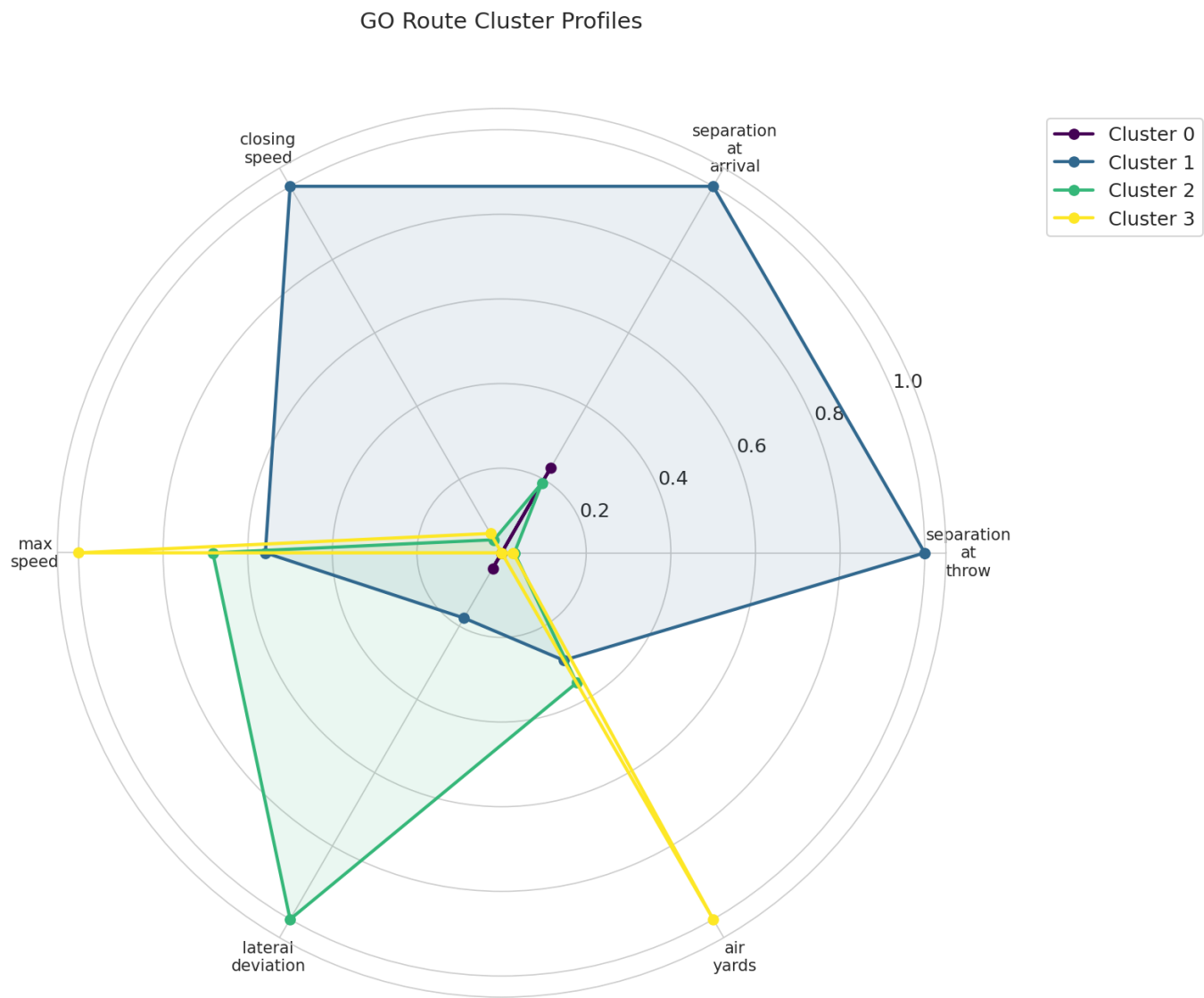


Question 2: Are There Distinct Types of GO Routes?

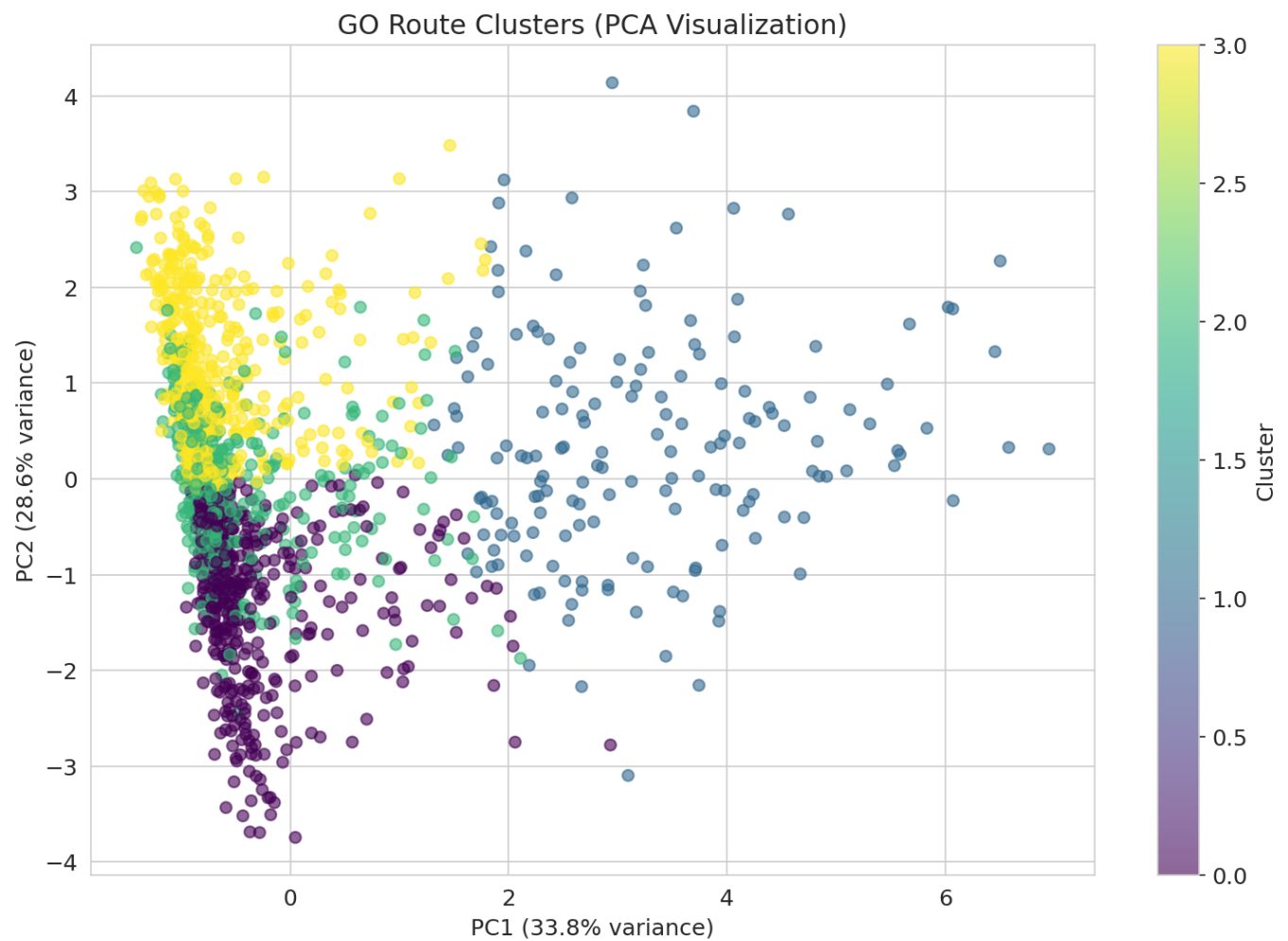
K-means clustering identified **4 distinct types** of GO routes:

Cluster	Profile	Completion Rate	Count
1	Wide Open - High separation (12 yds), fast closing	64%	167
0	Short & Tight - Short routes, tight coverage	45%	410

Cluster	Profile	Completion Rate	Count
2	High Movement - Significant lateral deviation	41%	341
3	Deep Bombs - Long air yards (35 yds), contested	35%	413



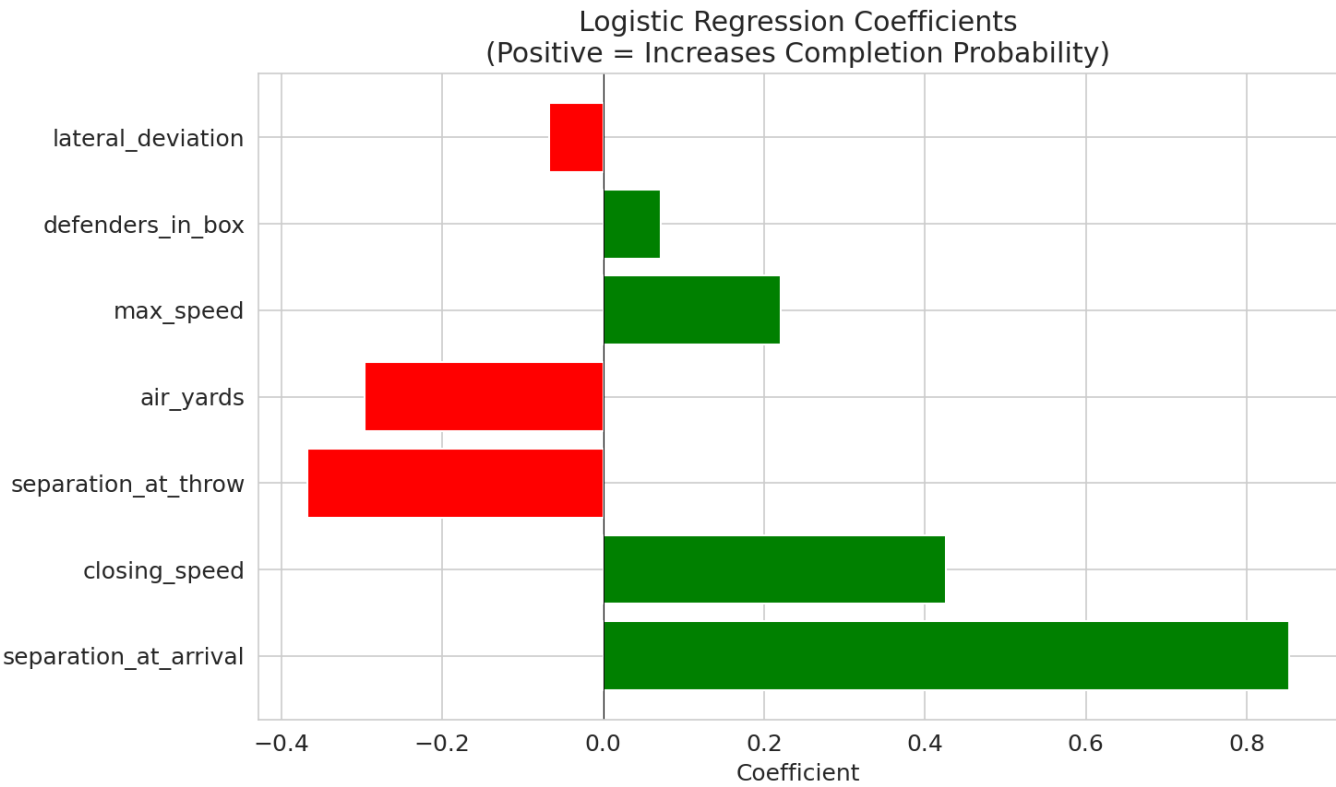
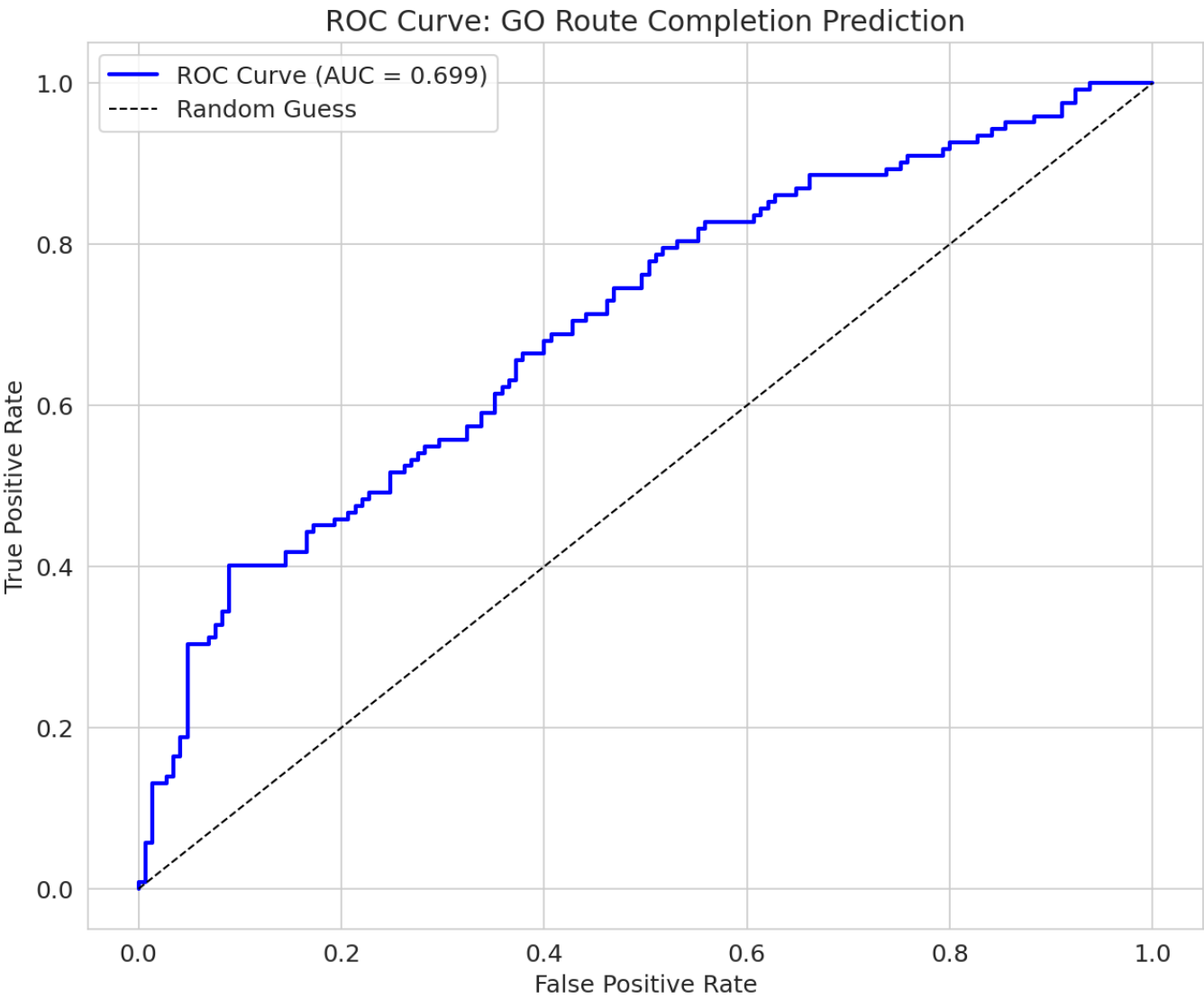
Key Insight: The "wide open" cluster (12 yards of separation at throw) has the highest success rate, but it's also the least common.



Question 3: Can We Predict Completion Probability?

Logistic Regression Model Performance: ROC-AUC = 0.699

Feature	Coefficient	Impact
Separation at Arrival	+0.85	Most important - more separation = higher completion
Closing Speed	+0.43	Higher closing speed correlates with completions
Separation at Throw	-0.37	Less important than arrival separation
Air Yards	-0.30	Longer throws = lower completion rate
Max Speed	+0.22	Slight positive effect
Defenders in Box	+0.07	Minimal impact
Lateral Deviation	-0.07	Minimal impact



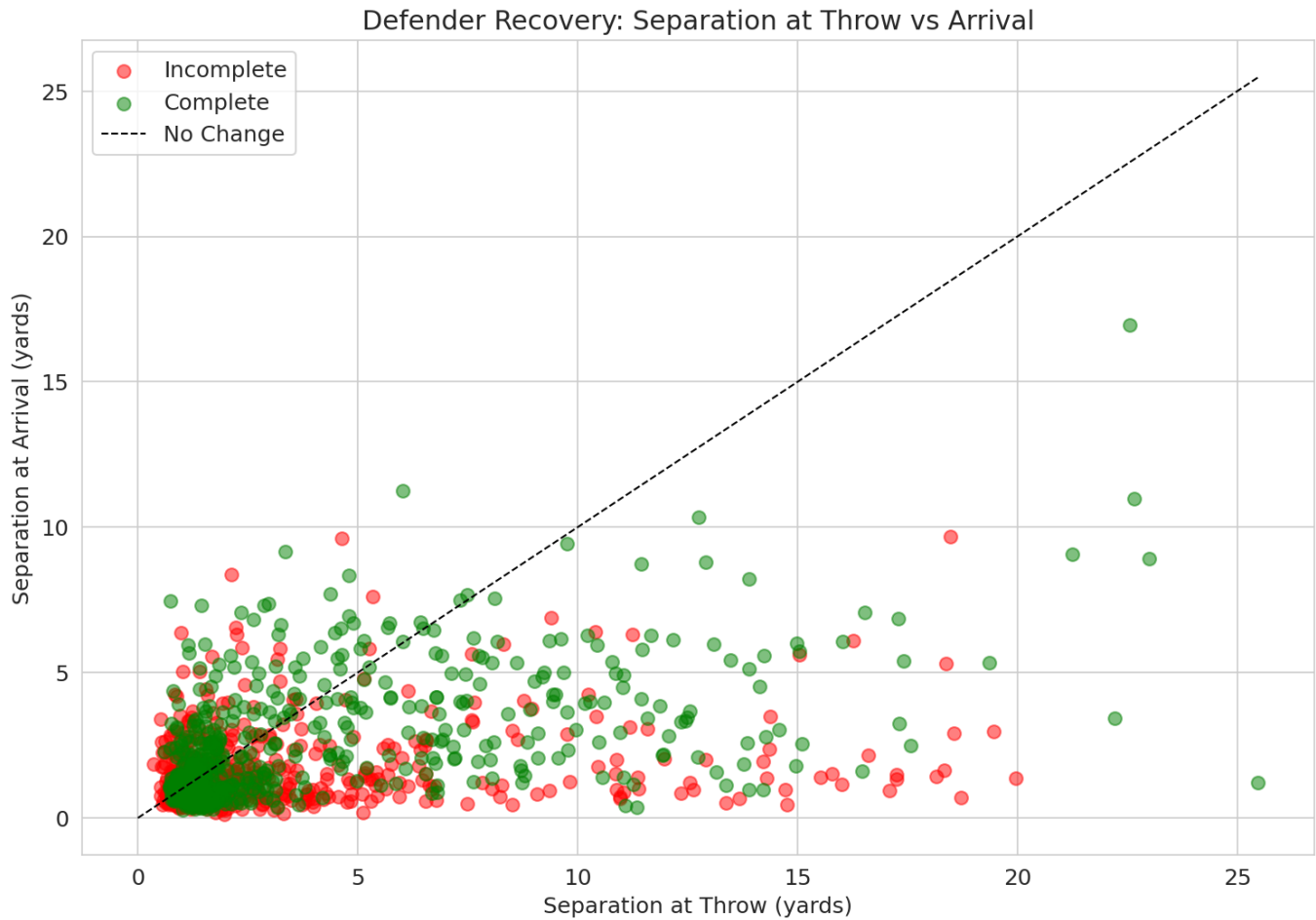
Key Insight: What matters most is where the receiver ends up (separation at arrival), not where they started.

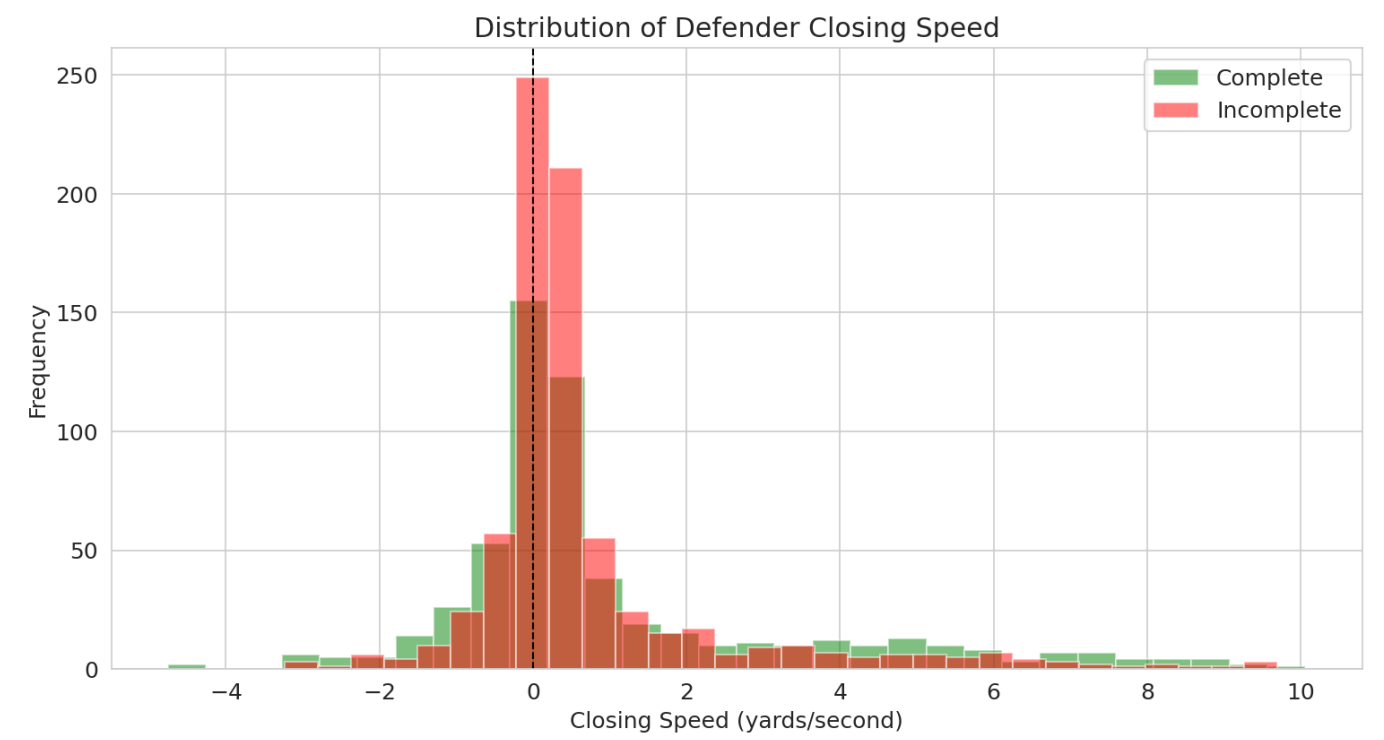
Question 4: How Do Defenders Recover?

Average closing speed: 0.84 yards/second

Outcome	Closing Speed	Sep at Throw	Sep at Arrival	Change
Incomplete	0.70 yd/s	2.78 yds	1.43 yds	-1.36 yds
Complete	1.03 yd/s	4.07 yds	2.51 yds	-1.55 yds

Counterintuitive Finding: Completions have *higher* closing speed. This suggests that when receivers get wide open, defenders make more aggressive recovery attempts, but it's too late.

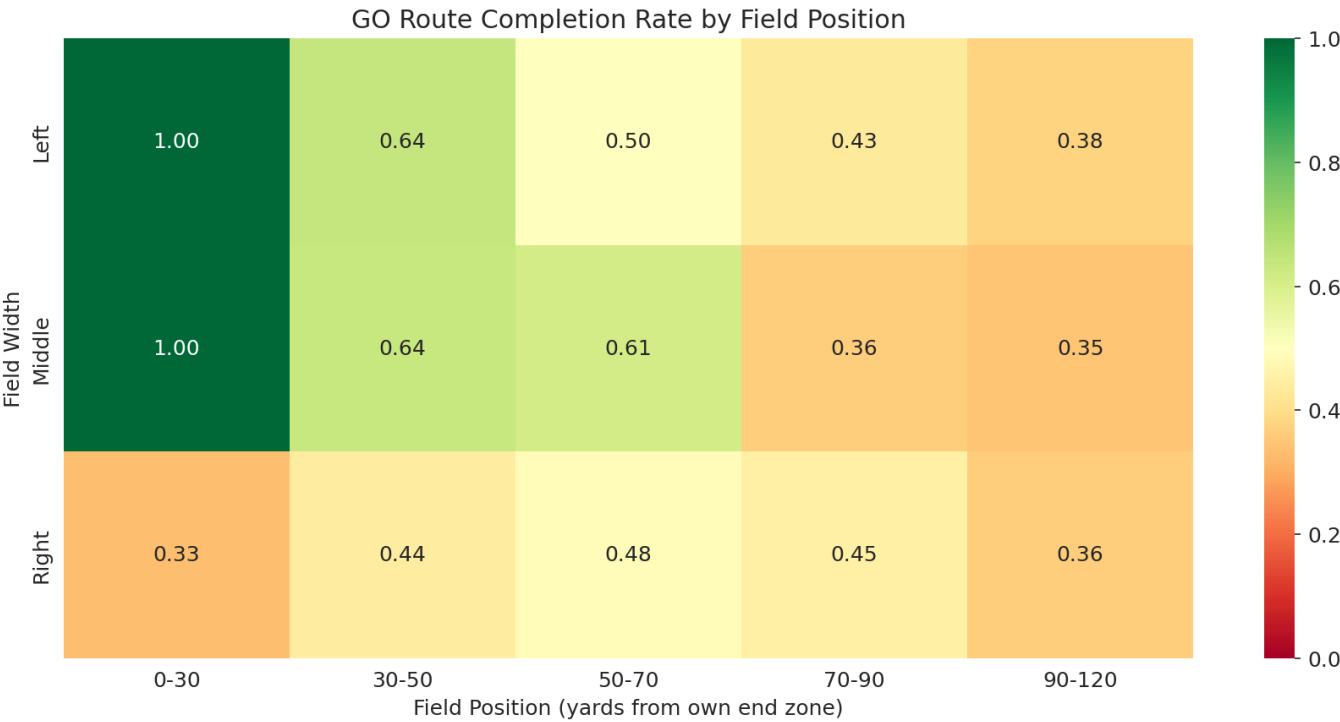




Question 5: Where Are GO Routes Most Effective?

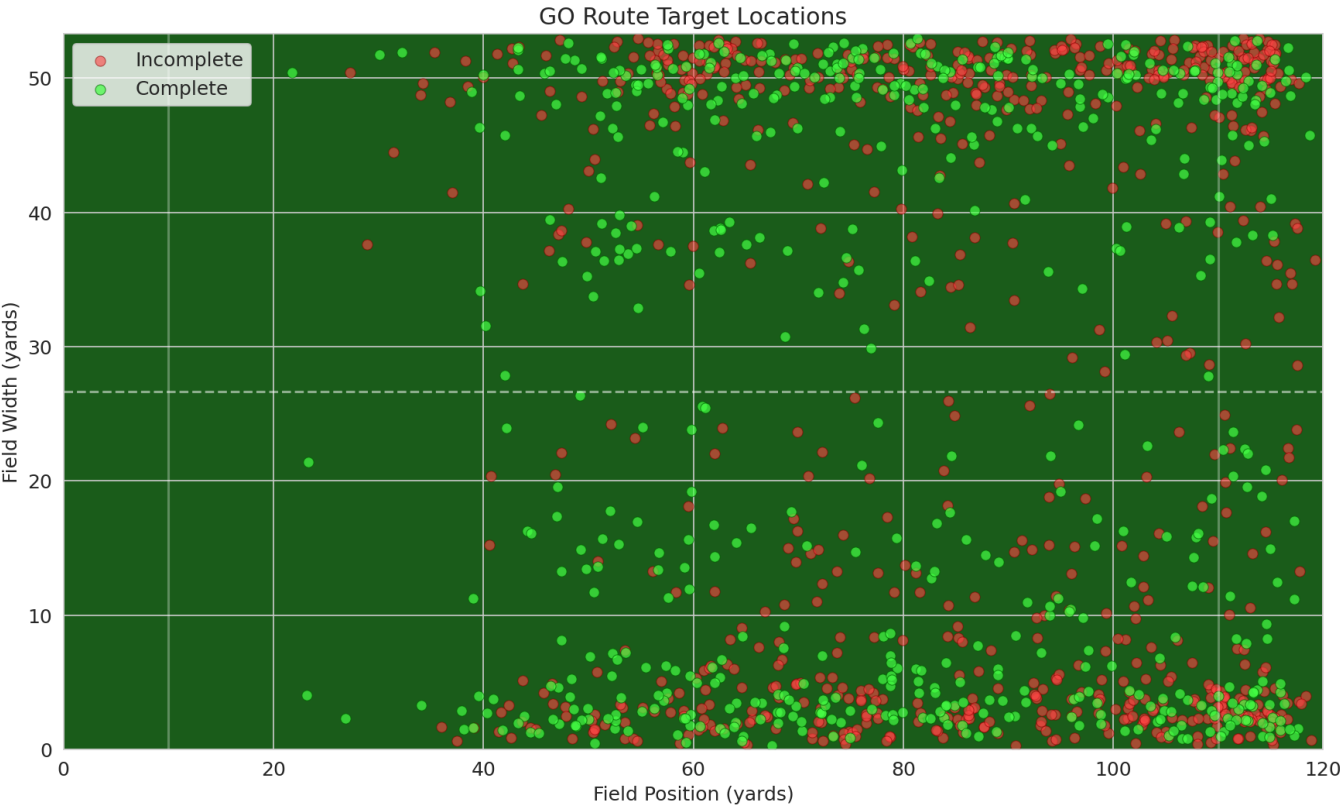
Hot Zones by Field Position:

Field Zone	Left	Middle	Right
30-50 yds	64%	64%	44%
50-70 yds	50%	61%	49%
70-90 yds	43%	36%	45%
90-120 yds	38%	35%	36%



Key Insights:

- Shorter GO routes (30-50 yards from own end zone) are most effective
- Middle of field offers slightly better odds
- Deep in opponent territory (90-120 yards) has lowest success rate (~35%)



Bonus: Man vs Zone Coverage

Coverage Type	Completion Rate	Count	Avg Separation	Closing Speed
---------------	-----------------	-------	----------------	---------------

Coverage Type	Completion Rate	Count	Avg Separation	Closing Speed
Man Coverage	40%	536	1.40 yds	0.31 yd/s
Zone Coverage	46%	795	2.23 yds	1.20 yd/s

Key Insight: GO routes are more successful against zone coverage (46% vs 40%), with 60% more separation at arrival.

Conclusions

For Offenses

- 1. **Target 2.5+ yards of separation at arrival** - this is the key threshold
- 2. **Zone coverage is exploitable** - GO routes succeed 46% vs 40% in man
- 3. **Shorter GO routes work better** - 30-50 yard field position has highest success
- 4. **The "wide open" route is rare but deadly** - when you get 12+ yards separation, take the shot

For Defenses

- 1. **Closing speed matters less than positioning** - don't get beat early
- 2. **Man coverage is better against GO routes** - but still only 60% success
- 3. **Contest at arrival** - separation at throw is less predictive than at arrival
- 4. **Deep in territory is favorable** - receivers succeed only 35% in red zone deep throws

Repository Structure

```
go_ball_portfolio/
├── README.md                # This file
├── go_ball_analysis.py      # Main analysis script
├── figures/                 # Visualizations
│   ├── q1_success_factors.png
│   ├── q2_cluster_pca.png
│   ├── q2_cluster_radar.png
│   ├── q2_elbow_curve.png
│   ├── q3_coefficients.png
│   ├── q3_roc_curve.png
│   ├── q4_closing_speed_dist.png
│   ├── q4_separation_scatter.png
│   ├── q5_field_scatter.png
│   └── q5_hot_zones_heatmap.png
└── results/                # Data exports
    ├── go_route_metrics.csv
    ├── summary_stats.csv
    ├── q1_statistical_tests.csv
    ├── q2_cluster_profiles.csv
    ├── q2_cluster_completion_rates.csv
    ├── q3_model_coefficients.csv
    └── q4_defender_recovery.csv
```

```
|— q5_hot_zones.csv  
|— bonus_coverage_analysis.csv
```

Technical Details

Data Processing

- Field direction standardized (all plays oriented left-to-right)
- Nearest defender tracked from arrival back to throw (Option A methodology)
- Time in air calculated from output frame count (10 fps)

Models

- **Clustering:** K-means with k=4 (selected via elbow method)
- **Prediction:** Logistic regression with standardized features
- **Statistical Tests:** Two-sample t-tests for metric comparisons

Tools

- Python (pandas, numpy, scikit-learn, matplotlib, seaborn)
- Kaggle Notebooks for computation

Author

Samuel Fraley Barcelona School of Economics - Data Science for Decision Making 2026 NFL Big Data Bowl

Links

- [Kaggle Competition](#)
- [NFL Next Gen Stats](#)