Do NFL Combine Metrics Actually Matter?

The NFL Combine plays a major role in player evaluation, but not all tests carry the same value across positions, so this project aimed to see whether NFL Combine tests meaningfully differentiate higher-level players within their own position groups. To do this, this analysis started by linking NFL Combine data with Pro Bowl selections as a proxy for sustained, high-level performance. Then test whether common Combine metrics separate Pro Bowl players from their peers by position. Pro Bowl status generally reflects a combination of on-field performance, and role stability relative to other players. The goal was to better understand whether Combine testing provides useful signal for evaluation.

Analytical Approach

For this analysis, NFL Combine data from 2000-2023 which was collected from Pro Football Reference. This included commonly used athletic testing metrics: 40 yard dash, vertical jump, bench press, broad jump, three-cone drill, shuttle drill, and body weight.

Pro Bowl selections were used to label players who reached a high performance tier within their position. This Pro Bowl status data was collected from Pro Football Reference over the same time period. Pro Bowl was measured as either a yes or no for each player across their career. Players were then grouped by position to account for different physical and athletic demands across roles.

Missing Combine values were excluded to maintain consistency across metrics, and summary statistics were first used to establish baseline differences in testing profiles between Pro Bowl and non-Pro Bowl players.

Statistical Methods

To evaluate whether Combine metrics meaningfully differentiate higher level players, one-way ANOVA test were applied within each position group. This approach allowed for comparison across multiple performance metrics while controlling for inflated error rates that would result from repeated pairwise testing. Quarterbacks were excluded from ANOVA testing due to incomplete Combine participation across metrics.

This analysis does not account for factors such as team context, injuries, and scheme, but logistic regression models were used to explore the relationship between combine performance and the probability of Pro Bowl selection within each position group. Pro Bowl status was treated as the binary outcome, with Combine metrics serving as predictors. Logistic regression was chosen over linear models to align with the binary nature of the outcome and to provide interpretable probability estimates rather than raw scores.

Results

Exploratory overview

Tabled results (appendix A) and visual comparisons (using density plots) (Appendix B) of Combine metrics consistently showed that Pro Bowl players often occupy different regions of the performance distributions than non-Pro Bowl players within the same position, though the degree of separation varied by both position and metric. In some cases, Pro Bowl players clustered toward the upper ends of specific metrics, while in others the distribution largely overlapped.

Across the full dataset, the empirical probability of a Combine participant reaching Pro Bowl status was approximately 8.5%, so essentially saying about 1 in every 12 athletes who were at the NFL Combine in years 2000-2023 achieved Pro Bowl status.

Position Specific Differences

When evaluated within position groups, several combine metrics showed consistent separation between Pro Bowl and non-Pro Bowl players. Measured via one-way ANOVA test, and the following is what Pro Bowl players performed better at than non-Pro Bowl players. Weight is translated as significance is heavier. Quarterbacks had insufficient participation in the Combine drills.

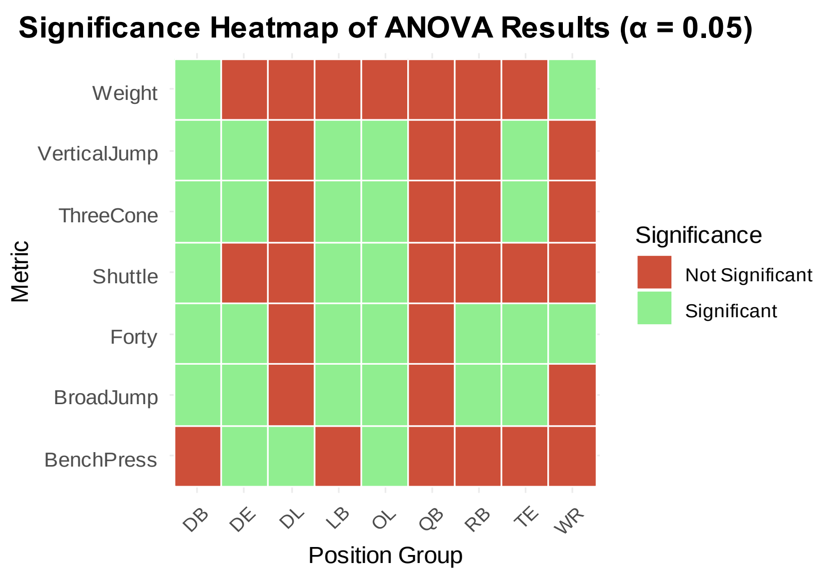

Figure 1. Shaded cells indicate metrics that significantly differentiated performance within a position group ($\alpha = 0.05$) ($p \leq 0.05$).

Positional Insights from ANOVA test

- **Defensive backs (DB):**
  Pro Bowl defensive backs consistently separated from their peers across multiple metrics. Differences in 40-yard dash, shuttle, three cone drill, vertical jump, and body weight. This suggests that high level DBs are efficient with their movement. Also heavier on average, suggesting that additional mass (larger frame, height/weight) may contribute to durability and sustained performance.
- **Offensive linemen (OL)**
  Pro Bowl offensive linemen showed separation across nearly all Combine metrics, including speed, power, and change of direction, while body weight did not differentiate outcomes. This

suggests that at comparable mass, higher level offensive linemen move better and bench more, rather than relying on size alone.

- **Defensive Linemen (DL)**
  Pro Bowl defensive linemen showed limited separation across most Combine metrics, with bench press performance emerging as the only significant differentiator. The lack of separation in body weight, speed, and explosiveness suggests that at comparable physical profiles, upper-body strength may play a more meaningful role in distinguishing higher-level interior defensive line performance.
- **Linebackers (LB)**
  Pro Bowl linebackers displayed stronger performances across most Combine metrics, excluding bench press. This suggests that higher level linebackers need to be more versatile/dynamic based on their ability to move faster, jump higher, and further.
- **Defensive Ends (DE)**
  Defensive ends demonstrated significant separation across all Combine metrics except the shuttle test. The lack of differentiation in body weight indicates that performance at this position is driven more by movement quality and explosiveness than by mass.
- **Running backs (RB)**
  Pro Bowl running backs separated in 40 yard dash performance and broad jump. This suggests that the key differentiators at running back are tied to acceleration and burst with increased lower body power from these metrics.
- **Wide receivers (WR)**
  Separation among Pro Bowl wide receivers was limited, with differences observed in the 40 yard dash, and body weight. This suggests that better wide receivers are bigger and faster, which may reflect the wide range of athletic profiles that can succeed at the position.
- **Tight ends (TE)**
  Pro Bowl tight ends displayed multiple differences in 40 yard dash, vertical jump, broad jump, and three cone drill. This potentially reflects the physical demands at this position to be able to have speed, power, and agility for blocking and receiving.

Supplementary Modeling Insights

Logistic regression models were applied within position groups to provide additional directional context for how Combine metrics relate to Pro Bowl outcomes. While these models were not intended as predictive tools, their results generally aligned with the exploratory findings.

Across many positions, greater body mass showed a positive association with Pro Bowl selection, reinforcing the importance of size and physicality in NFL roles. Strength and power-related measures further contributed directional signal for interior positions, while speed and agility metrics showed more nuanced and position specific effects.

Overall, the regression results served as a consistency check, supporting the conclusion that Combine metrics provide contexts that must be interpreted relative to position demands.

Takeaways

This analysis started with my exploratory approach to see if the NFL combine helps distinguish higher level players. By comparing Pro Bowl and non-Pro Bowl players within position groups, the analysis shows that there may be some informative takeaways dependent on the position.

For some roles, such as offensive linemen, defensive end, linebacker, and defensive backs, combine metrics consistently separated higher level players across multiple athletic domains. In this case Pro Bowl as the success metric, success was often tied to movement efficiency and power at similar body mass rather than size alone. For other positions including wide receiver and interior defensive linemen, Combine testing showed more limited separation, suggesting that these tests alone may underrepresent performance relevant traits,.

Overall, the results show that the Combine works best when it's viewed through a position specific lens. Instead of using the same benchmarks for everyone. This information could be used to better weight testing results and support team specific evaluating measures.

Ultimately, the value of the NFL Combine lies not in universal benchmarks, but in understanding how athletic traits match the demands of each position, and where those traits meaningfully separate players competing for the same roles. This exploratory framework can be extended by incorporating draft position, or role-specific usage data to further contextualize how athletic traits translate to on-field performance

Appendix A

POSITION	*WEIGHT (LBS)*	*40-YD (SEC)*	*VERTICAL JUMP (INCH)*	*BENCH PRESS (REPS: 225)*	*BROAD JUMP (INCH)*	*THREE CONE (SEC)*	*SHUTTLE (SEC)*
DB	198.6	4.5	35.8	15.5	121.7	7.0	4.2
DE	266.1	4.8	33.1	23.7	115.5	7.3	4.4
DL	303.6	5.1	29.6	27.2	106.3	7.7	4.6
LB	239.8	4.7	34.1	22.2	117.3	7.1	4.3
OL	312.6	5.3	28.0	25.2	102.4	7.8	4.7
QB	221.9	4.8	31.2	19.3	111.1	7.2	4.3
RB	213.4	4.6	34.4	19.3	118.4	7.1	4.3
TE	253.3	4.8	33.1	20.2	115.4	7.2	4.4
WR	201.6	4.5	35.4	14.6	121.1	7.0	4.2

Table A1: Table shows the average value of NFL Combine metrics at each position group years: 2000-2023**.**

POSITION	*WEIGHT (LBS)*	*40-YD (SEC)*	*VERTICAL JUMP (INCH)*	*BENCH PRESS (REPS: 225)*	*BROAD JUMP (INCH)*	*THREE CONE (SEC)*	*SHUTTLE (SEC)*
DB	202.2	4.5	36.5	15.5	123.4	6.9	4.2
DE	268.7	4.7	34.6	25.2	117.9	7.2	4.4
DL	308.4	5.0	29.9	30.3	107.5	7.6	4.6
LB	241.6	4.6	35.8	23.0	120.0	7.0	4.2
OL	312.2	5.1	29.2	27.5	105.4	7.7	4.7
QB	223.3	4.8	31.3	23.0	113.0	7.1	4.3
RB	215.4	4.5	34.9	20.2	120.4	7.0	4.3
TE	255.1	4.7	34.6	21.2	118.9	7.1	4.4
WR	205.1	4.5	36.0	15.5	121.5	7.0	4.2

Table A2: Table shows the average value of NFL Combine metrics at each position group for Pro Bowl players, years: 2000-2023

Appendix B

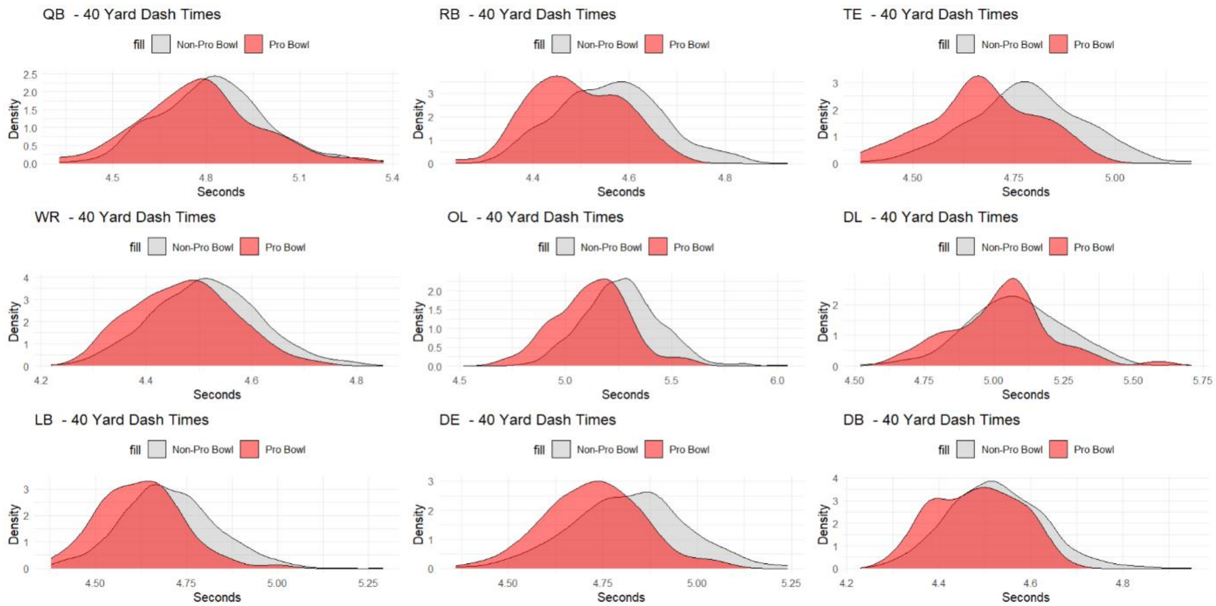

Figure B1: Density Plot distribution of 40 Yard Dash performance by position. NFL Combine average vs Pro Bowl player average.

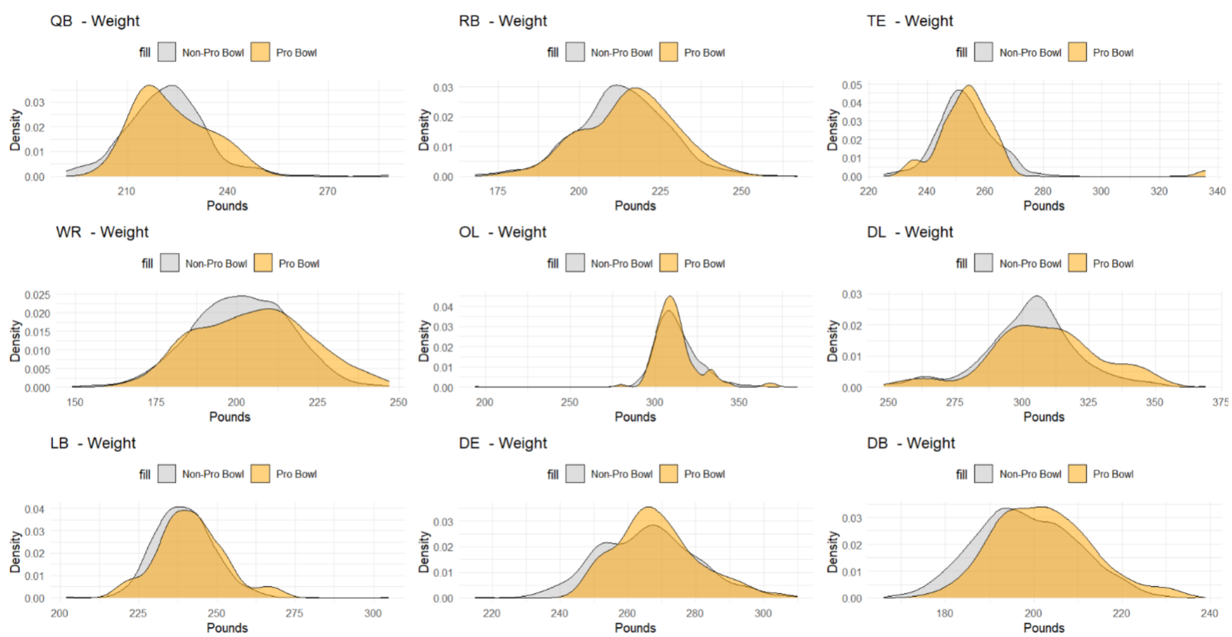

Figure B2: Density Plot comparing Weight by position. NFL Combine average vs Pro Bowl player average.

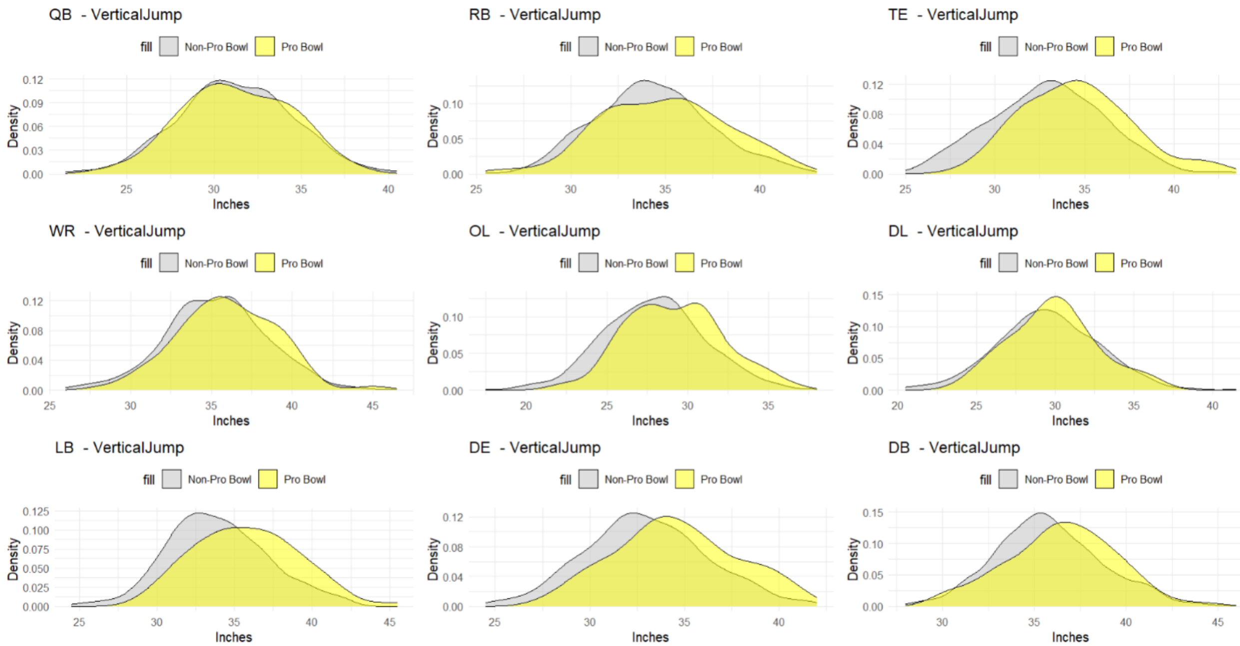

Figure B3: Density Plot comparing Vertical Jump performance by position. NFL Combine average vs Pro Bowl player average.

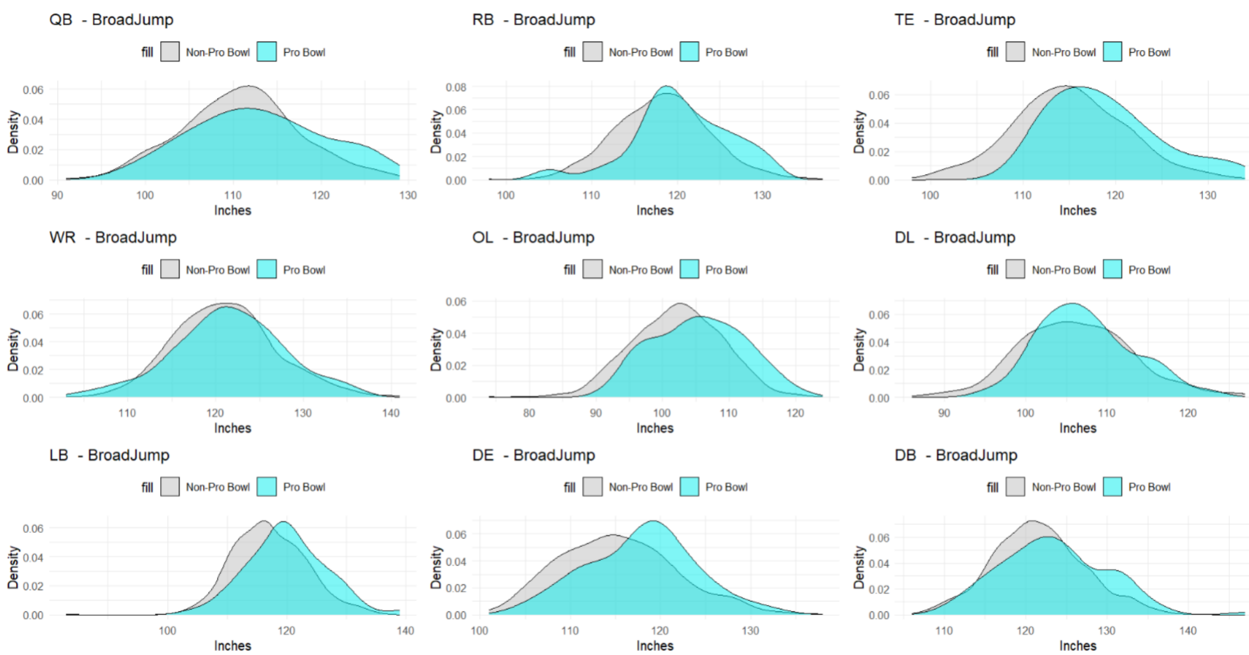

Figure B4: Density Plot comparing Broad Jump performance by position. NFL Combine average vs Pro Bowl player average**.**

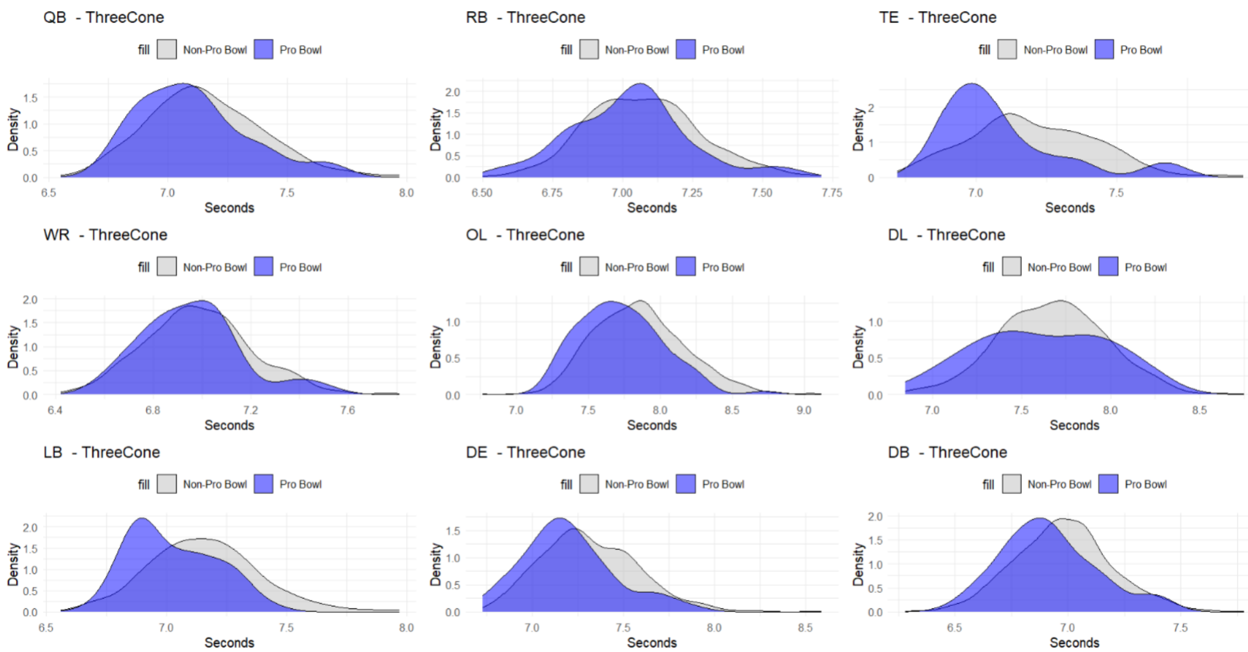

Figure B5: Density Plot comparing Three Cone performance by position. NFL Combine average vs Pro Bowl player average.

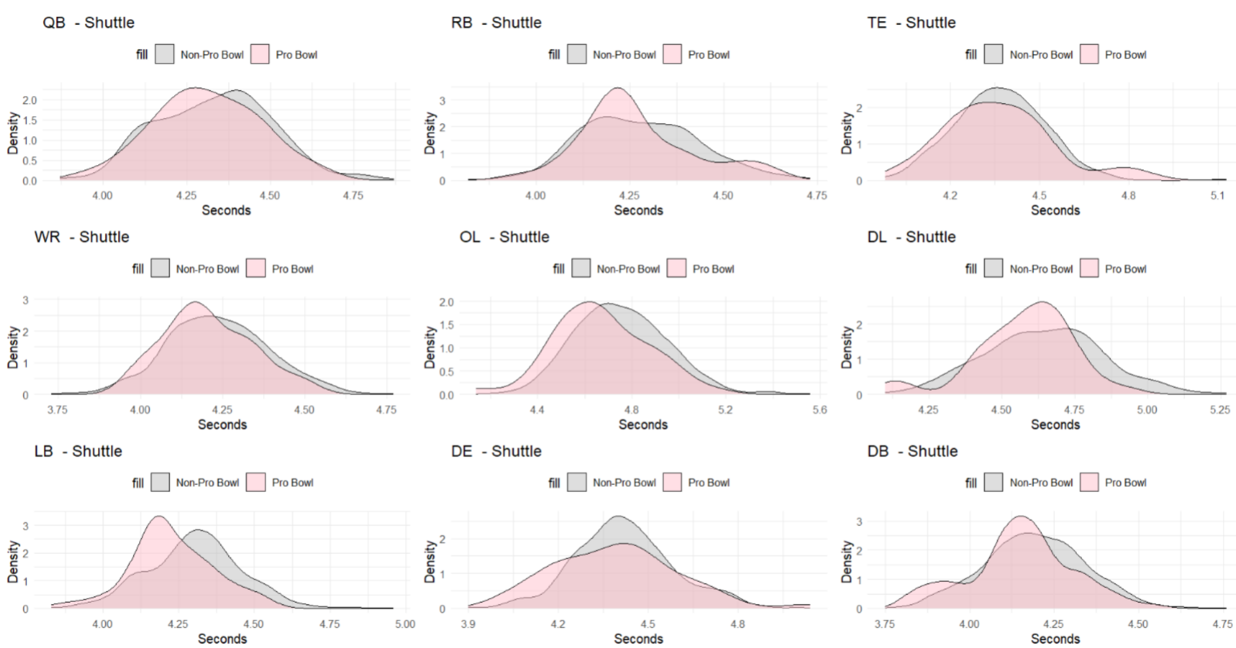

Figure B6: Density Plot comparing Shuttle performance by position. NFL Combine average vs Pro Bowl player average.

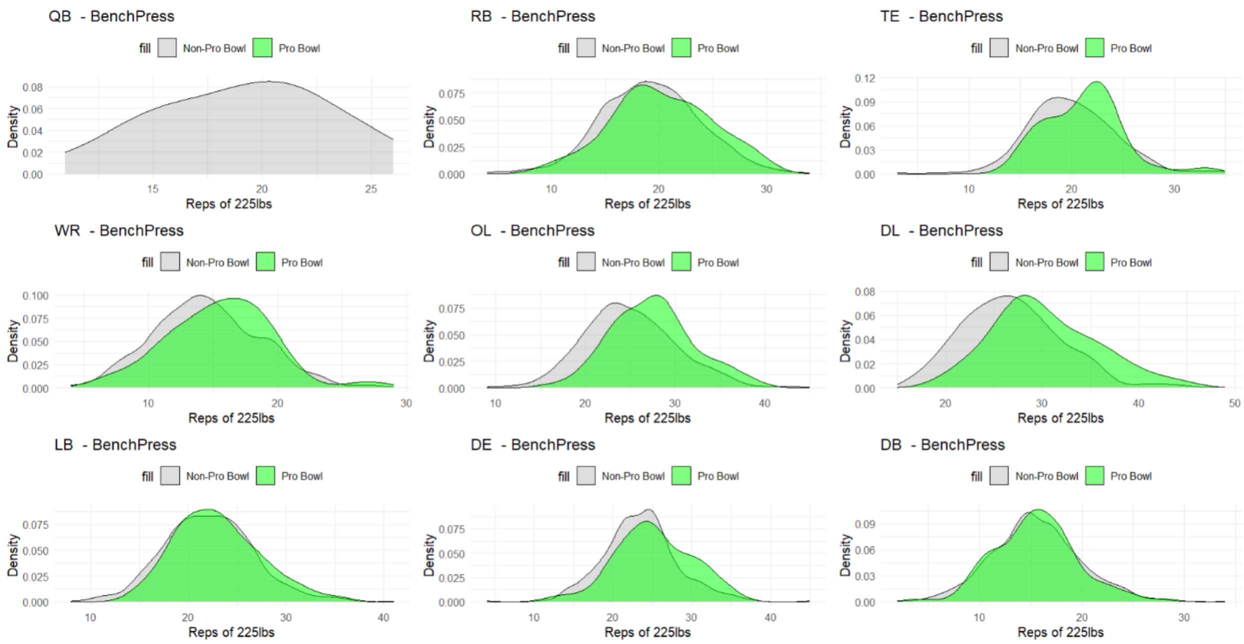

Figure B7: Density Plot comparing Bench Press performance by position. NFL Combine average vs Pro Bowl player average.

Appendix C

RB Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	23.0941	14.0044	1.6491	10706414858.1973	0.0991	
WEIGHT	0.0372	0.0175	2.1265	1.0379	0.0335	*
FORTY	-8.6118	2.4436	-3.5242	0.0002	0.0004	***
VERTICALJUMP	-0.0510	0.0916	-0.5565	0.9503	0.5779	
BENCHPRESS	-0.0062	0.0448	-0.1394	0.9938	0.8892	
BROADJUMP	0.0591	0.0492	1.2006	1.0609	0.2299	
THREECONE	-0.8646	1.2394	-0.6976	0.4212	0.4854	
SHUTTLE	1.5550	1.6838	0.9235	4.7350	0.3557	

Significance Levels: *** $p < 0.001$ **$p < 0.01$ *$p < 0.05$
Goodness-of-fit (measure): *McFadden* R^2 *= 0.0114 ~ 11.4%*

OL Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	7.2780	7.7658	0.9372	1448.1043	0.3487	
WEIGHT	0.0329	0.0126	2.6131	1.0334	0.0090	**
FORTY	-3.2239	1.0350	-3.1150	0.0398	0.0018	**
VERTICALJUMP	0.0120	0.0552	0.2183	1.0121	0.8272	
BENCHPRESS	0.0490	0.0270	1.8129	1.0502	0.0698	
BROADJUMP	0.0316	0.0275	1.1498	1.0321	0.2502	
THREECONE	-0.4776	0.6767	-0.7059	0.6202	0.4803	
SHUTTLE	-0.9019	1.0384	-0.8685	0.4058	0.3851	

Significance Levels: *** $p < 0.001$ **$p < 0.01$ *$p < 0.05$
Goodness-of-fit (measure): *McFadden* R^2 *= 0.109 ~ 10.9%*

TE Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	13.4847	16.3989	0.8223	########	0.4109	
WEIGHT	0.0721	0.0264	2.7303	1.0747	0.0063	**
FORTY	-7.4187	2.6004	-2.8529	0.0006	0.0043	**
VERTICALJUMP	0.0383	0.0981	0.3906	1.0390	0.6961	
BENCHPRESS	-0.0280	0.0662	-0.4232	0.9724	0.6721	
BROADJUMP	0.0181	0.0577	0.3130	1.0182	0.7543	
THREECONE	-1.5237	1.4950	-1.0192	0.2179	0.3081	
SHUTTLE	2.0363	1.9069	1.0679	7.6621	0.2856	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.1645 ~ 16.5%*

WR Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	20.0955	15.3342	1.3105	533768910.6931	0.1900	
WEIGHT	0.0073	0.0181	0.4061	1.0074	0.6847	
FORTY	-4.4559	2.6958	-1.6529	0.0116	0.0984	
VERTICALJUMP	0.0588	0.0922	0.6374	1.0605	0.5238	
BENCHPRESS	0.0289	0.0544	0.5313	1.0293	0.5952	
BROADJUMP	-0.0157	0.0496	-0.3160	0.9844	0.7520	
THREECONE	0.6535	1.3312	0.4909	1.9223	0.6235	
SHUTTLE	-2.2105	1.8313	-1.2071	0.1097	0.2274	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.0478 ~ 4.8%*

DL Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	4.6782	13.4687	0.3473	107.5775	0.7283	
WEIGHT	0.0404	0.0183	2.2074	1.0413	0.0273	*
FORTY	-3.4068	1.9140	-1.7799	0.0331	0.0751	
VERTICALJUMP	-0.1606	0.0997	-1.6105	0.8516	0.1073	
BENCHPRESS	0.0874	0.0430	2.0332	1.0913	0.0420	*
BROADJUMP	0.0590	0.0531	1.1110	1.0607	0.2666	
THREECONE	0.3052	1.1485	0.2657	1.3569	0.7904	
SHUTTLE	-1.8800	1.7686	-1.0629	0.1526	0.2878	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.1068 ~ 10.7%*

DE Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	17.4770	14.7671	1.1835	38918184.1833	0.2366	
WEIGHT	0.0838	0.0206	4.0718	1.0874	0.0000	***
FORTY	-6.4958	2.0746	-3.1311	0.0015	0.0017	**
VERTICALJUMP	0.0754	0.0907	0.8313	1.0783	0.4058	
BENCHPRESS	-0.0375	0.0451	-0.8298	0.9632	0.4067	
BROADJUMP	-0.0201	0.0479	-0.4197	0.9801	0.6747	
THREECONE	-3.2087	1.2459	-2.5753	0.0404	0.0100	*
SHUTTLE	2.8717	1.6309	1.7608	17.6674	0.0783	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.1811 ~ 18.1%*

LB Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	4.8403	13.1262	0.3687	126.5023	0.7123	
WEIGHT	0.0661	0.0208	3.1774	1.0683	0.0015	**
FORTY	-2.6409	1.8784	-1.4060	0.0713	0.1597	
VERTICALJUMP	0.1863	0.0748	2.4902	1.2048	0.0128	*
BENCHPRESS	-0.0065	0.0408	-0.1603	0.9935	0.8726	
BROADJUMP	0.0277	0.0409	0.6787	1.0281	0.4973	
THREECONE	-2.6876	1.1262	-2.3865	0.0680	0.0170	*
SHUTTLE	-0.3580	1.5629	-0.2291	0.6991	0.8188	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.19661 ~ 19.7%*

DB Logistic Regression Model

METRIC	ESTIMATE	STDERROR	ZVALUE	ODDSRATIO	PVALUE	SIGNIFICANCE
(INTERCEPT)	14.5343	10.0585	1.4450	2051851.5177	0.1485	
WEIGHT	0.0463	0.0139	3.3348	1.0474	0.0009	***
FORTY	-4.4441	1.6141	-2.7533	0.0117	0.0059	**
VERTICALJUMP	-0.0508	0.0616	-0.8241	0.9505	0.4099	
BENCHPRESS	-0.0417	0.0357	-1.1686	0.9592	0.2426	
BROADJUMP	0.0444	0.0335	1.3262	1.0454	0.1848	
THREECONE	-1.1574	0.8053	-1.4372	0.3143	0.1506	
SHUTTLE	-0.2275	1.1163	-0.2038	0.7965	0.8385	

Significance Levels: **** $p < 0.001$ **$p < 0.01$ *$p < 0.05$*
Goodness-of-fit (measure): *McFadden R^2 = 0.0622 ~ 6.2%*

Appendix D

The logistic regression models in this appendix are intended to provide directional context rather than predictive classification. Each model estimates how NFL Combine metrics relate to the likelihood of achieving Pro Bowl status within a given position group.

- **Estimate:** Indicates the direction of the relationship between a Combine metric and Pro Bowl selection. Positive values suggest higher odds, while negative values suggest lower odds.
- **Odds Ratio:** Represents the multiplicative change in odds for a one-unit increase in the metric. For example, an odds ratio of 1.05 indicates a 5% increase in odds, while 0.95 indicates a 5% decrease.
- **Statistical Significance:** Metrics meeting conventional significance thresholds ($p < 0.05$) suggest a consistent association within the sample. Non-significant results do not imply irrelevance, but rather limited evidence of separation given the available data.
- **Model Fit:** McFadden's R^2 values are reported to provide context on explanatory power. These values are typically lower than linear R^2 and should be interpreted comparatively across positions rather than as absolute performance measures.

These models should be interpreted alongside the ANOVA results and distributional analyses. Together, they provide complementary perspectives on how athletic traits relate to higher-level performance within position groups.