

SCOUTING COMBINE





THE PROBLEM

- How can the NFL teams successfully predict who will get taken each day of the draft
- Combine provides teams an opportunity to see how athletically gifted a player is
- What is the difference between a drafted and an undrafted player





THE DATA

- **Draft information, combine testing, and college stats web scraped from sportsreference.com**
- **Additional data taken from the `nfl_data_py` package**
- **Every combine and draft since 2005**





SUMMARY

- Combine performance importance
- Easier to distinguish drafted from undrafted than days 1-2 from day 3



	height	weight	speed	shuttle_agility	cone_agility	vertical	bench_reps	broad_jump	forty_yd	cone	shuttle
Round											
0.0	73.503244	237.049954	88.401341	123.765949	122.881392	32.091985	19.259958	113.540159	4.815307	7.299085	4.440916
1.0	73.771672	243.699811	94.601681	131.525983	128.218943	32.847725	20.975228	114.865316	4.761854	7.252012	4.399943
2.0	74.157716	246.776145	99.821009	136.640225	132.612171	33.672163	21.771523	117.143781	4.713309	7.216081	4.376118

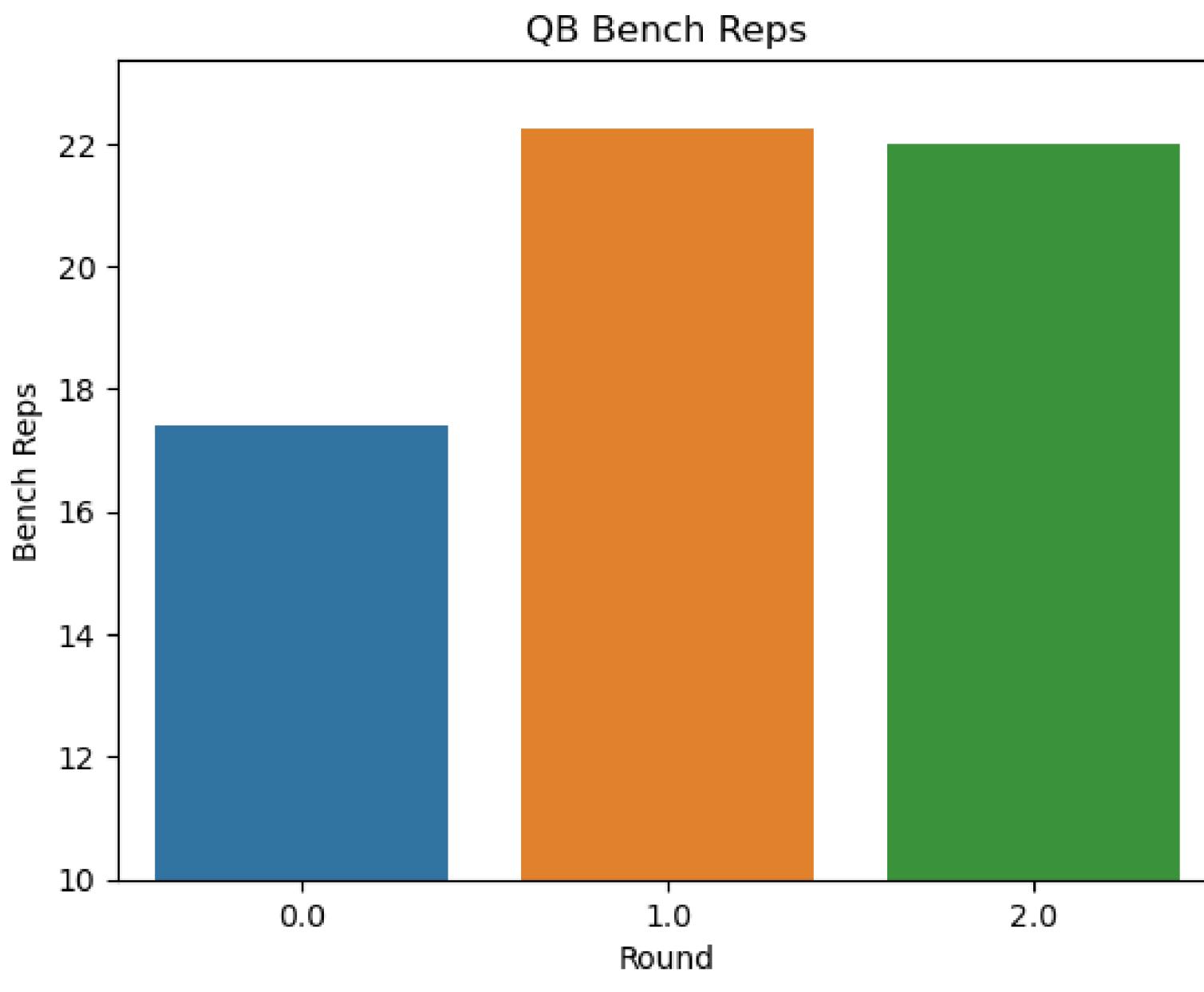
Top Correlations with Round



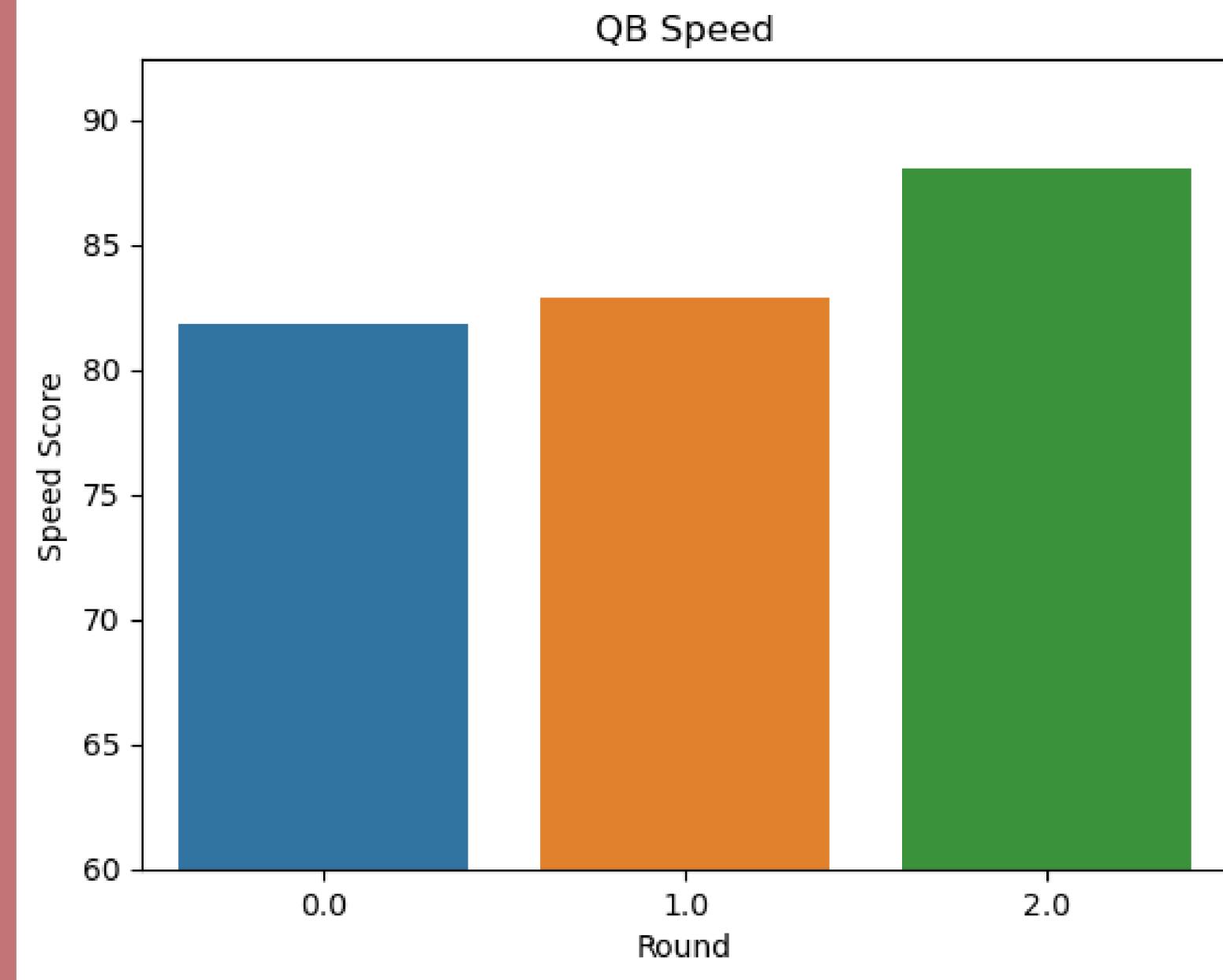


QB SUMMARY

QB Bench Reps



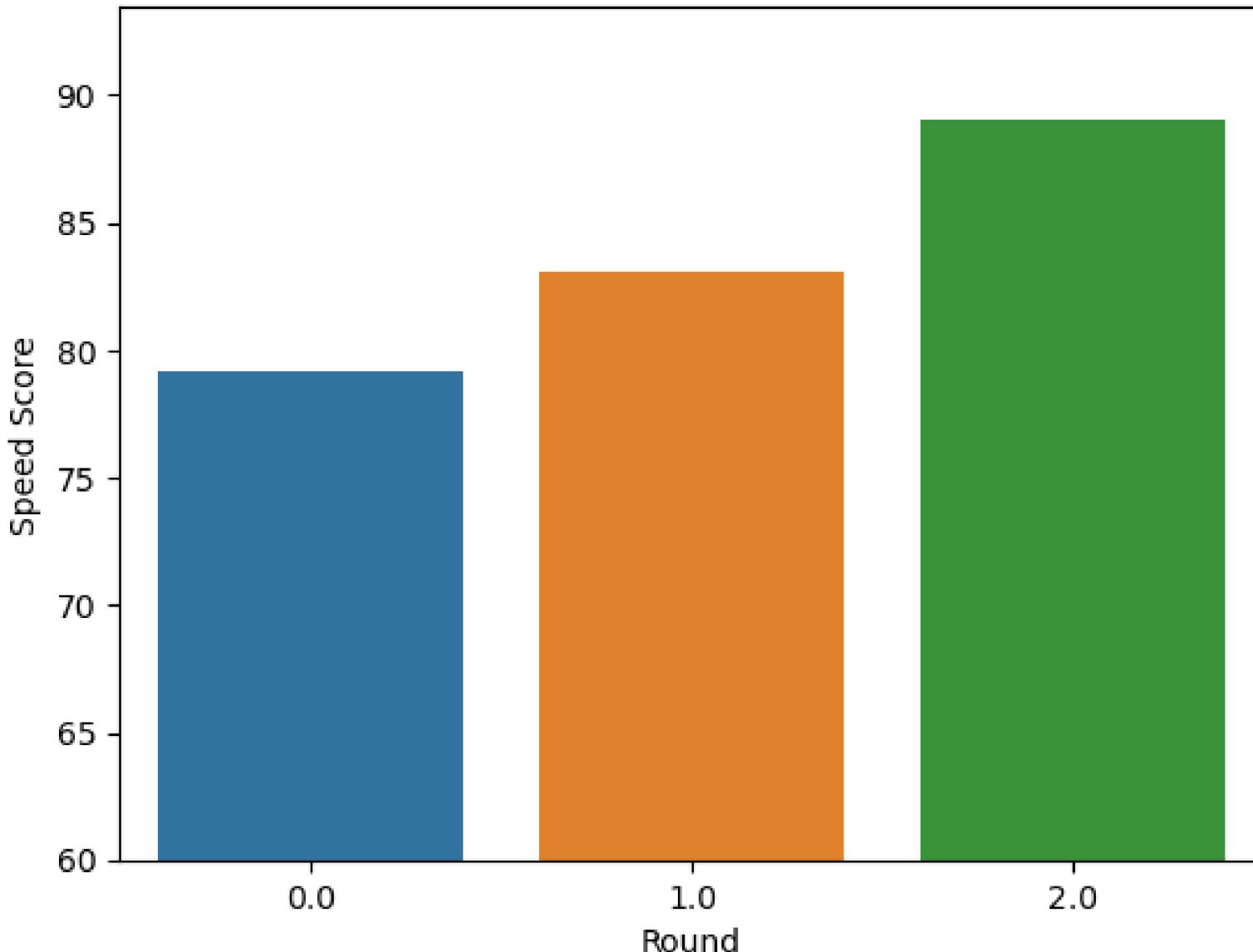
QB Speed



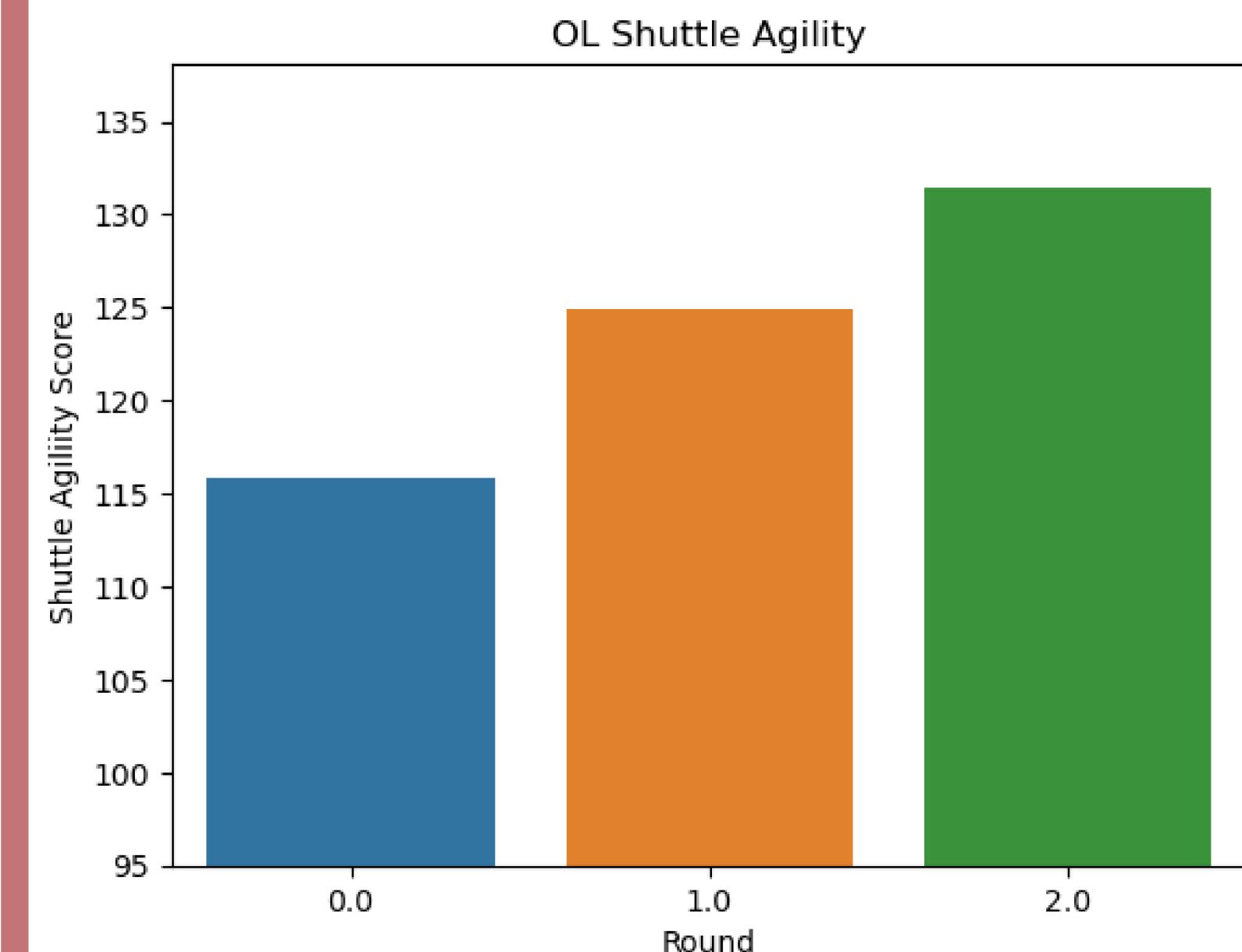


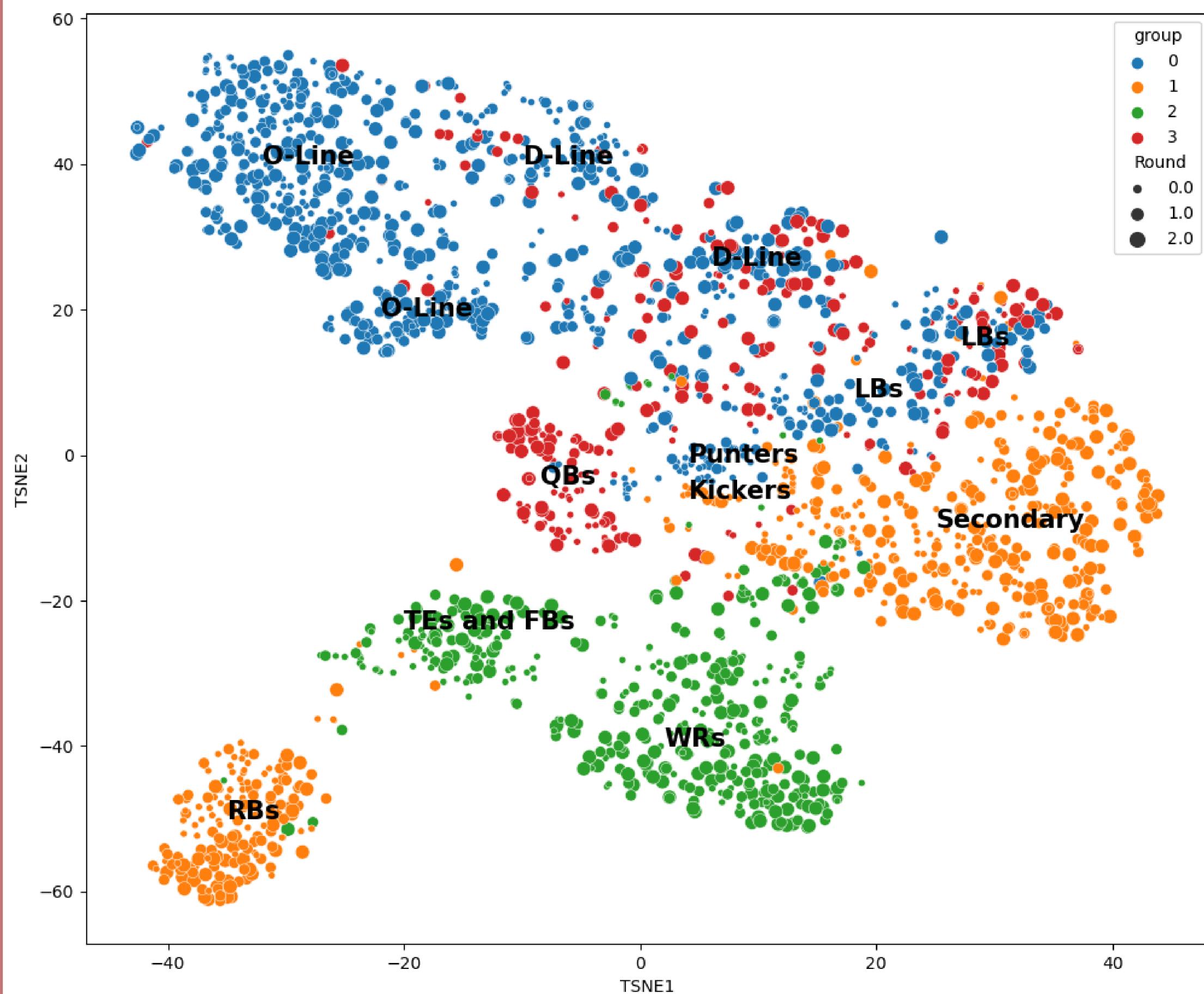
OL SUMMARY

OL Speed



OL Shuttle Agility







MODELS

Random Forest Model

	precision	recall	f1-score	support
0.0	0.73	0.88	0.80	717
1.0	0.54	0.45	0.49	695
2.0	0.59	0.54	0.56	591
accuracy			0.63	2003
macro avg	0.62	0.63	0.62	2003
weighted avg	0.62	0.63	0.62	2003

Decision Tree Model

	precision	recall	f1-score	support
0.0	0.0	0.70	0.83	0.76
1.0	1.0	0.49	0.43	0.46
2.0	2.0	0.52	0.49	0.51
accuracy			0.59	2003
macro avg	0.57	0.58	0.58	2003
weighted avg	0.58	0.59	0.58	2003

XGBoost Model

	precision	recall	f1-score	support
0.0	0.65	0.97	0.78	717
1.0	0.53	0.38	0.44	695
2.0	0.53	0.39	0.45	591
accuracy			0.60	2003
macro avg	0.57	0.58	0.56	2003
weighted avg	0.57	0.60	0.57	2003

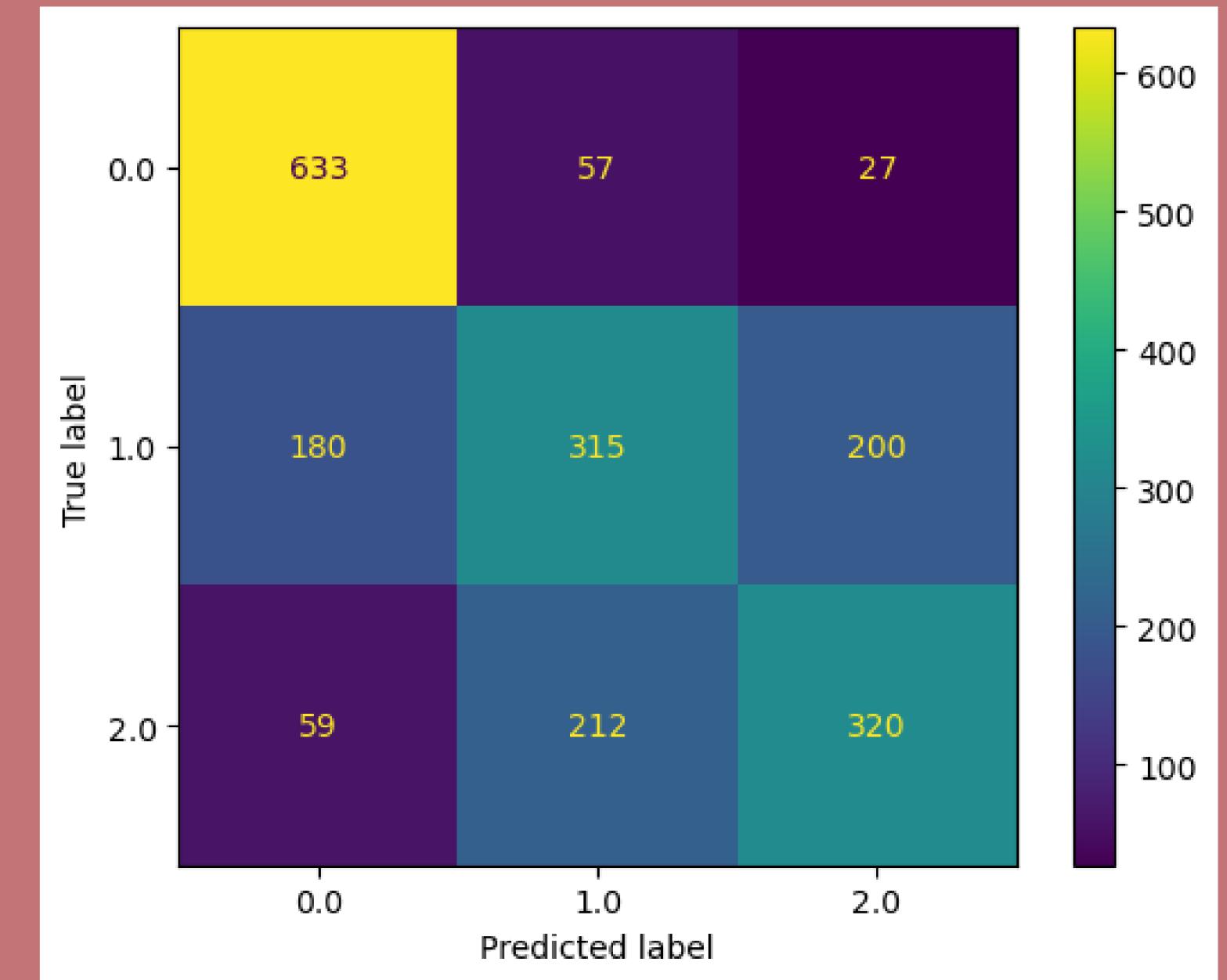
KNN

	precision	recall	f1-score	support
0.0	0.0	0.58	0.86	0.69
1.0	1.0	0.46	0.31	0.37
2.0	2.0	0.58	0.45	0.51
accuracy			0.55	2003
macro avg	0.54	0.54	0.52	2003
weighted avg	0.54	0.55	0.53	2003



BEST MODEL

- **Best Model: Random Forest**
- **Parameters:**
 - **Criterion: Gini**
 - **Max Depth: None**
 - **Min Sample Leaf: 2**
 - **Min Sample Split: 2**
 - **N_Estimators: 80**
 - **SMOTE**





TOP FEATURES

Column	Coefficient
Cone_Agility	0.12
DL	0.04
DE	0.04
Solo Tackles Season	0.03
EDGE	0.03
DT	0.028



CONCLUSIONS

- Combine athletic testing is very significant in separating the drafted from the undrafted
- Combination of athleticism and physical attributes
- College offensive stats have more meaning than defensive stats
- Get more numerical data
 - Advanced stats
 - Non Primary Stats
 - More bio information
 - Spend more time feature engineering
- Add scouting report information and other non numeric data





QUESTIONS?

- Josh Palgon
- linkedin.com/in/josh-palgon/
- <https://github.com/jpalgon>
- jopalgon@gmail.com