

Assembling A Dream Team Using NBA Player Archetypes

The Dream Team (Makers)

Jasmine Huynh

Brian Rivers

Matt Melby

Chris Lewicki



Project Overview

- Research datasets and clean them
- Conduct exploratory data analysis (EDA)
- Perform feature engineering
- Build and assess the machine learning models

NBA Regular Season Stats 23-24:

<https://www.kaggle.com/datasets/vivovinco/2023-2024-nba-player-stats?resource=download&select=2023-2024+NBA+Player+Stats+-+Regular.csv>

NBA Playoffs Season Stats 23-24:

<https://www.kaggle.com/datasets/vivovinco/2023-2024-nba-player-stats?resource=download&select=2023-2024+NBA+Player+Stats+-+Playoffs.csv>

NBA Salaries 23-24:

<https://hoopshype.com/salaries/players/2023-2024/>

Regular Season Dataset

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	1	Precious Achiuwa	PF-C	24	TOT	74	18	21.9	3.2	6.3	...	0.616	2.6	4.0	6.6	1.3	0.6	0.9	1.1	1.9	7.6
1	1	Precious Achiuwa	C	24	TOR	25	0	17.5	3.1	6.8	...	0.571	2.0	3.4	5.4	1.8	0.6	0.5	1.2	1.6	7.7
2	1	Precious Achiuwa	PF	24	NYK	49	18	24.2	3.2	6.1	...	0.643	2.9	4.3	7.2	1.1	0.6	1.1	1.1	2.1	7.6
3	2	Bam Adebayo	C	26	MIA	71	71	34.0	7.5	14.3	...	0.755	2.2	8.1	10.4	3.9	1.1	0.9	2.3	2.2	19.3
4	3	Ochai Agbaji	SG	23	TOT	78	28	21.0	2.3	5.6	...	0.661	0.9	1.8	2.8	1.1	0.6	0.6	0.8	1.5	5.8
...
730	568	Thaddeus Young	PF	35	PHO	10	0	8.9	1.1	2.1	...	0.333	1.7	1.1	2.8	0.7	0.5	0.2	0.4	1.1	2.3
731	569	Trae Young	PG	25	ATL	54	54	36.0	8.0	18.7	...	0.855	0.4	2.3	2.8	10.8	1.3	0.2	4.4	2.0	25.7
732	570	Omer Yurtseven	C	25	UTA	48	12	11.4	2.1	3.8	...	0.679	1.5	2.8	4.3	0.6	0.2	0.4	0.8	1.1	4.6
733	571	Cody Zeller	C	31	NOP	43	0	7.4	0.6	1.4	...	0.605	1.1	1.5	2.6	0.9	0.2	0.1	0.4	1.0	1.8
734	572	Ivica Zubac	C	26	LAC	68	68	26.4	5.0	7.6	...	0.723	2.9	6.3	9.2	1.4	0.3	1.2	1.2	2.6	11.7

735 rows x 30 columns

Playoffs Season Dataset

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	1	Precious Achiuwa	PF	24	NYK	9	2	20.4	2.3	4.8	...	0.385	2.4	1.8	4.2	0.6	0.4	1.3	0.4	2.2	5.2
1	2	Bam Adebayo	C	26	MIA	5	5	38.4	9.6	19.4	...	0.714	2.2	7.2	9.4	3.8	0.4	0	1.6	3	22.6
2	3	Nickeil Alexander-Walker	SG	25	MIN	16	1	23.6	2.6	7	...	1	0.3	1.4	1.8	2.3	0.6	0.4	0.9	1.5	7.3
3	4	Grayson Allen	SG	28	PHO	2	2	21.5	0.5	2.5	...	1	1	3	4	1	0.5	0	1.5	2	3.5
4	5	Jarrett Allen	C	25	CLE	4	4	31.8	6.3	9.3	...	0.692	3.8	10	13.8	1.3	1.3	1	1.3	2	17
...
209	210	Jaylin Williams	C	21	OKC	10	0	12.7	1.6	3.3	...	0.75	0.4	2.8	3.2	1.5	0.4	0.3	0.3	1.1	4.4
210	211	Kenrich Williams	PF	29	OKC	7	0	4.6	0.3	1.1	...	0	0.1	1	1.1	0.4	0.1	0	0	0.4	0.6
211	212	Delon Wright	PG	31	MIA	4	1	26.8	3	5	...	1	0.5	2.5	3	1.8	1	0.3	0	1.3	8
212	213	Thaddeus Young	PF	35	PHO	1	0	4	0	0	...	0	0	0	0	0	0	0	0	0	0
213	214	Ivica Zubac	C	26	LAC	6	6	32	7	11.7	...	0.65	3	6.3	9.3	1	0.8	0.5	1.3	3.5	16.2

214 rows × 30 columns

NBA Salary Dataset

	Rank	Player	Salary_2023_24	Salary_2023_24_Alternate
1	1.	Stephen Curry	\$51,915,615	\$51,915,615
2	2.	Kevin Durant	\$47,649,433	\$47,649,433
3	3.	Nikola Jokic	\$47,607,350	\$47,607,350
4	3.	LeBron James	\$47,607,350	\$47,607,350
5	3.	Joel Embiid	\$47,607,350	\$47,607,350
...
608	607.	Jalen Crutcher	\$64,343	\$64,343
609	607.	Izaiah Brockington	\$64,343	\$64,343
610	607.	Taevion Kinsey	\$64,343	\$64,343
611	611.	Edmond Sumner	\$40,459	\$40,459
612	612.	Kaiser Gates	\$35,389	\$35,389

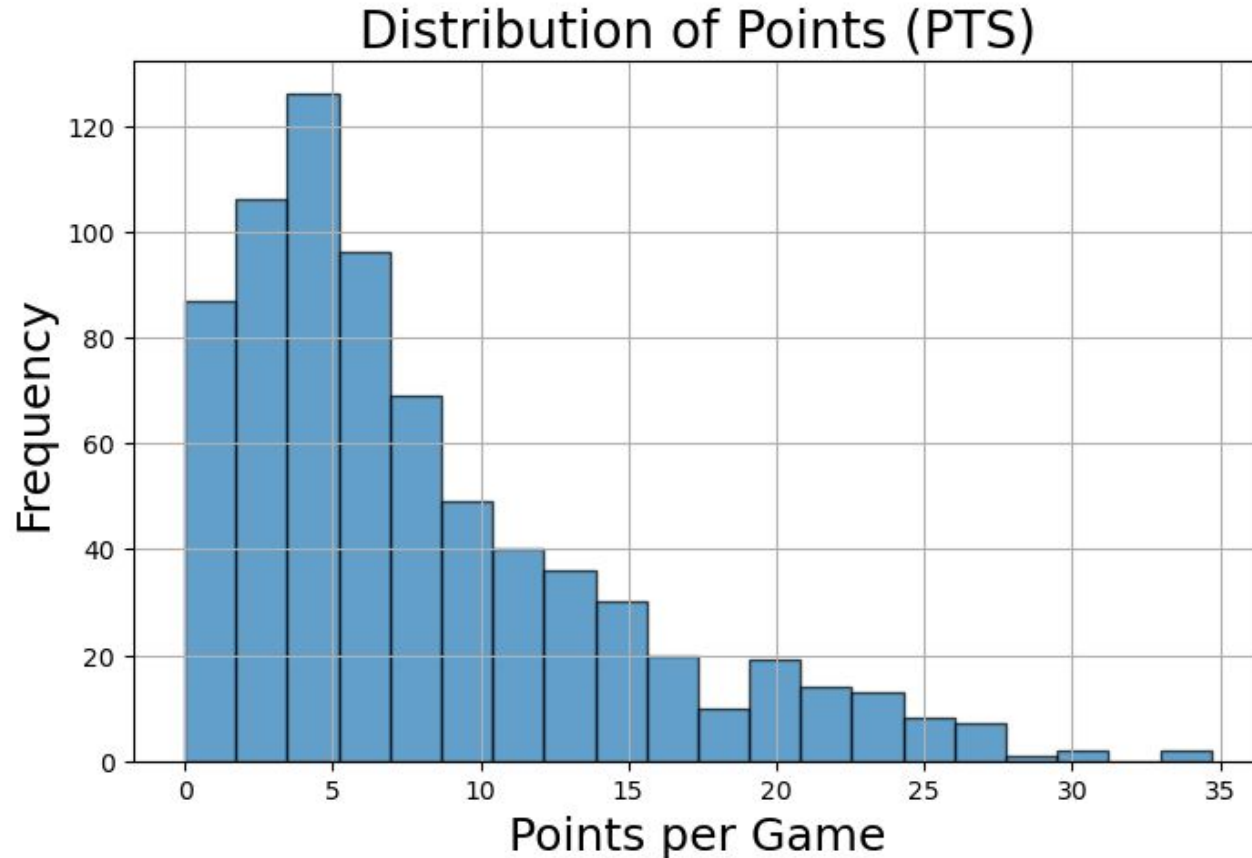
612 rows × 4 columns

Exploratory Data Analysis

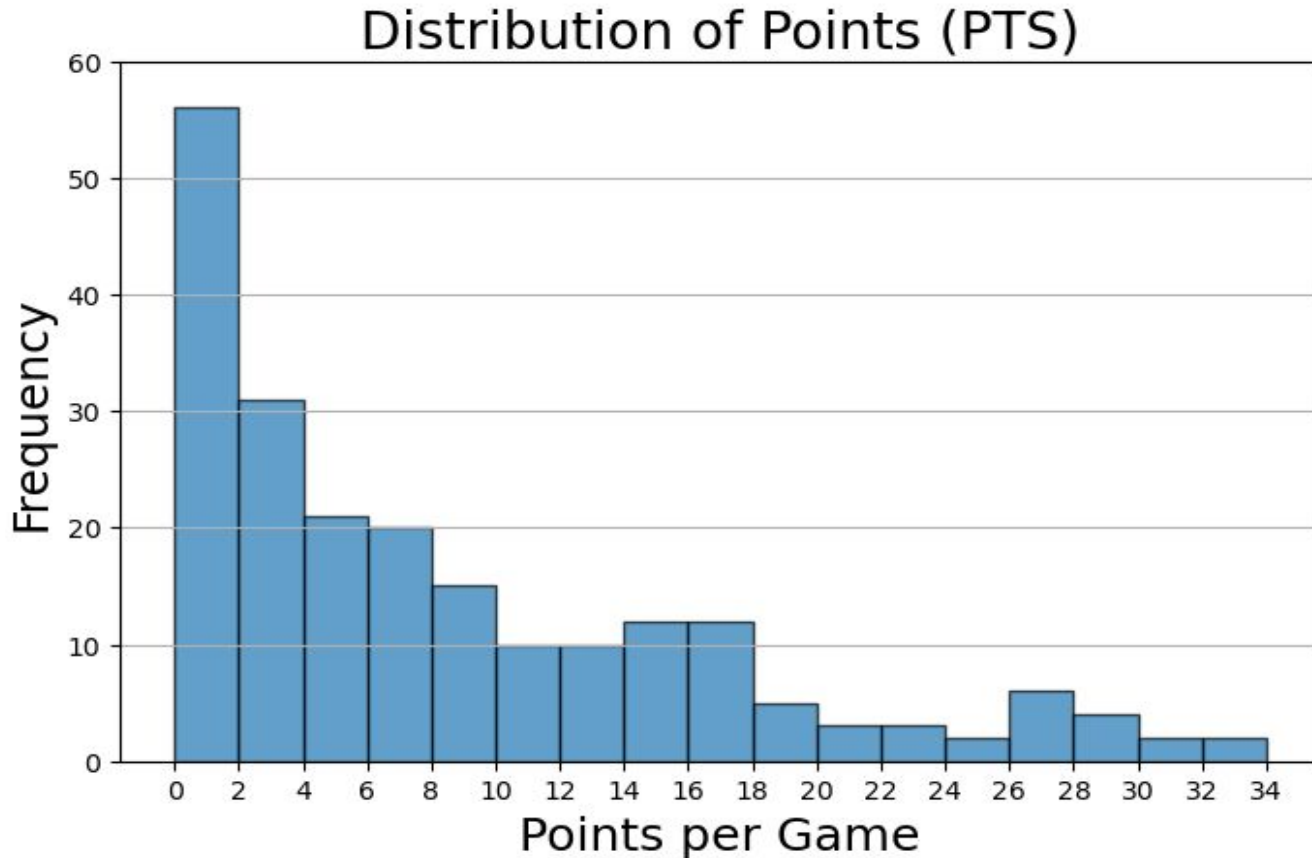
During this phase, we focused on uncovering key patterns and trends in the 2023–2024 NBA player statistics dataset.

The goal was to identify the most significant features influencing player performance and roles to guide our clustering and archetype definition.

Scoring Distribution During Regular Season (PTS)

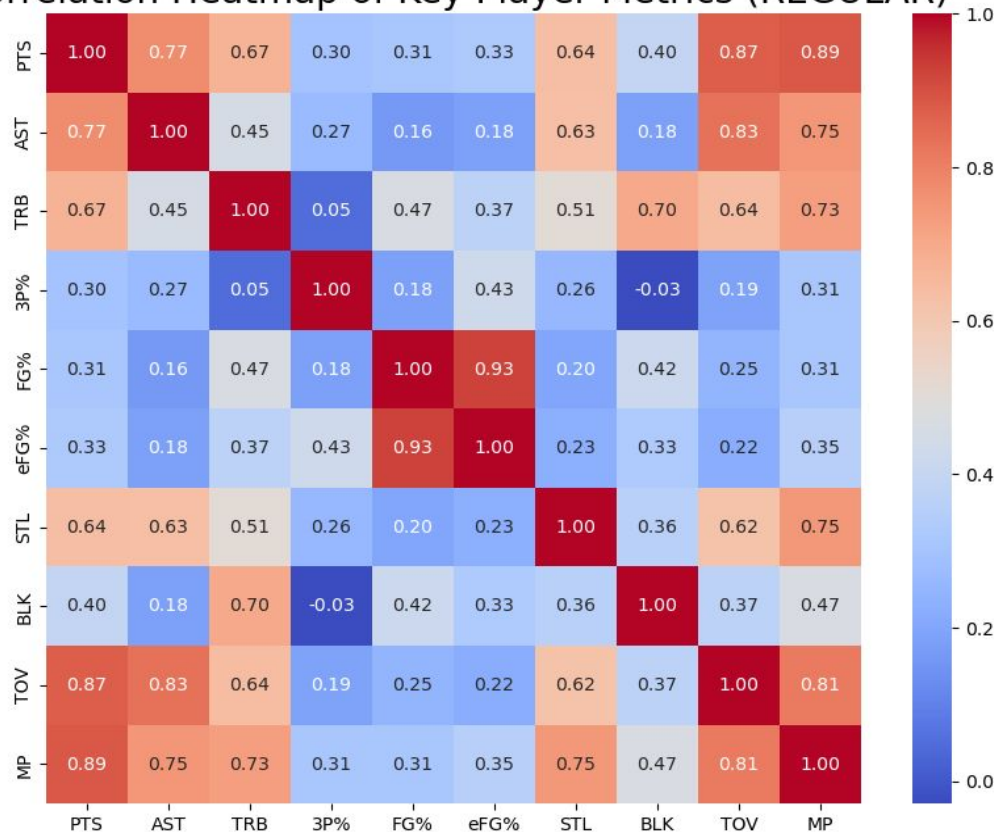


Scoring Distribution During Playoffs (PTS)



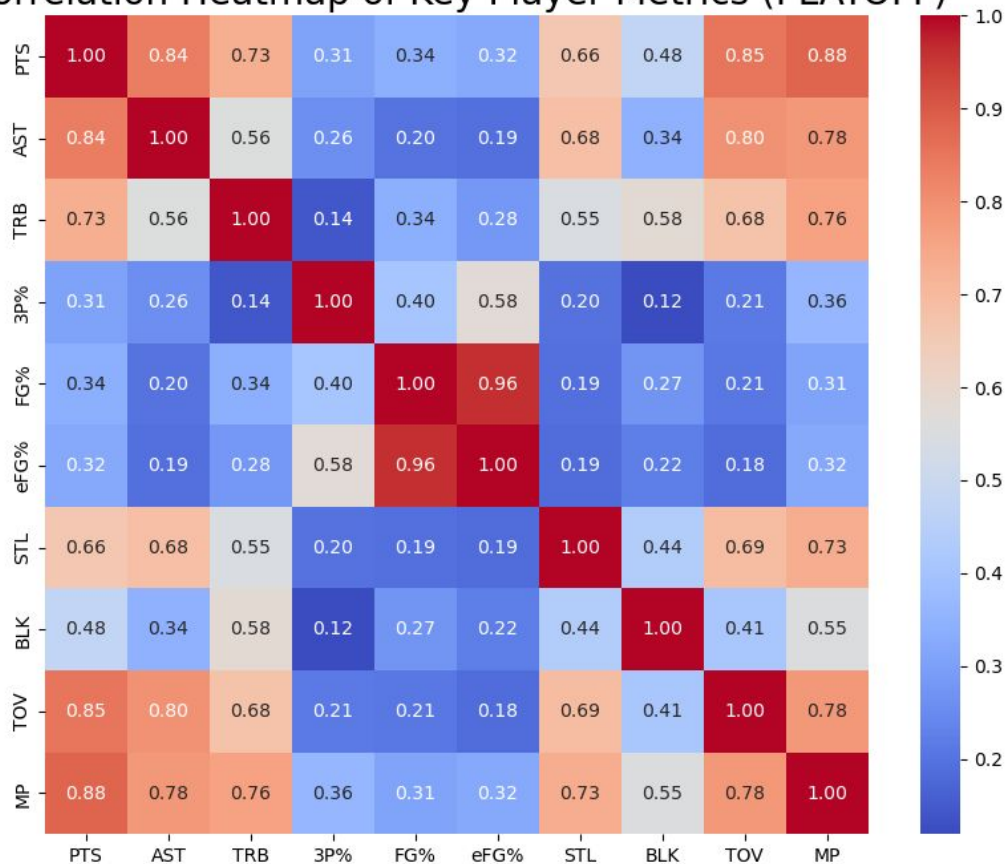
Key Correlations Heatmap (Regular Season)

Correlation Heatmap of Key Player Metrics (REGULAR)

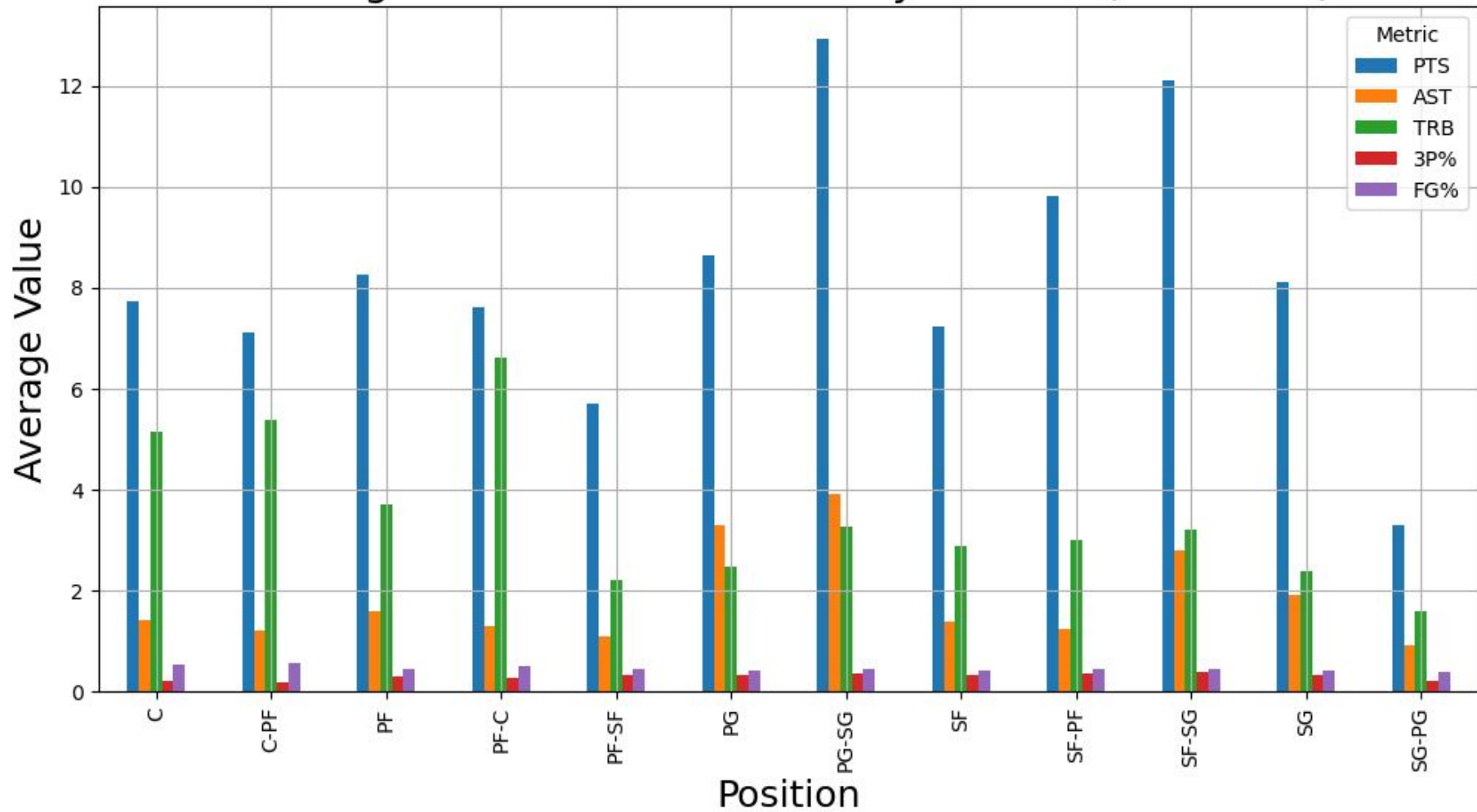


Key Correlations Heatmap (Playoffs)

Correlation Heatmap of Key Player Metrics (PLAYOFF)

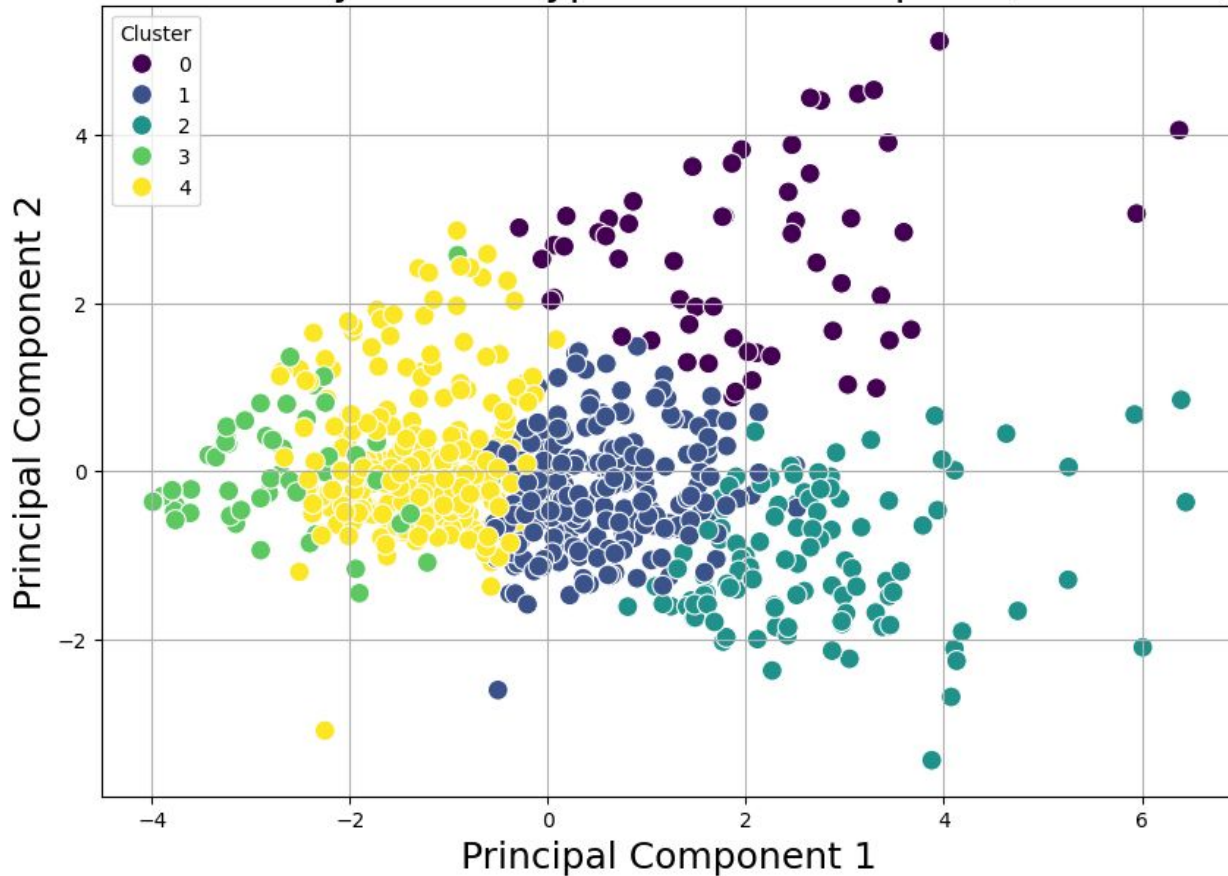


Average Performance Metrics by Position (REGULAR)

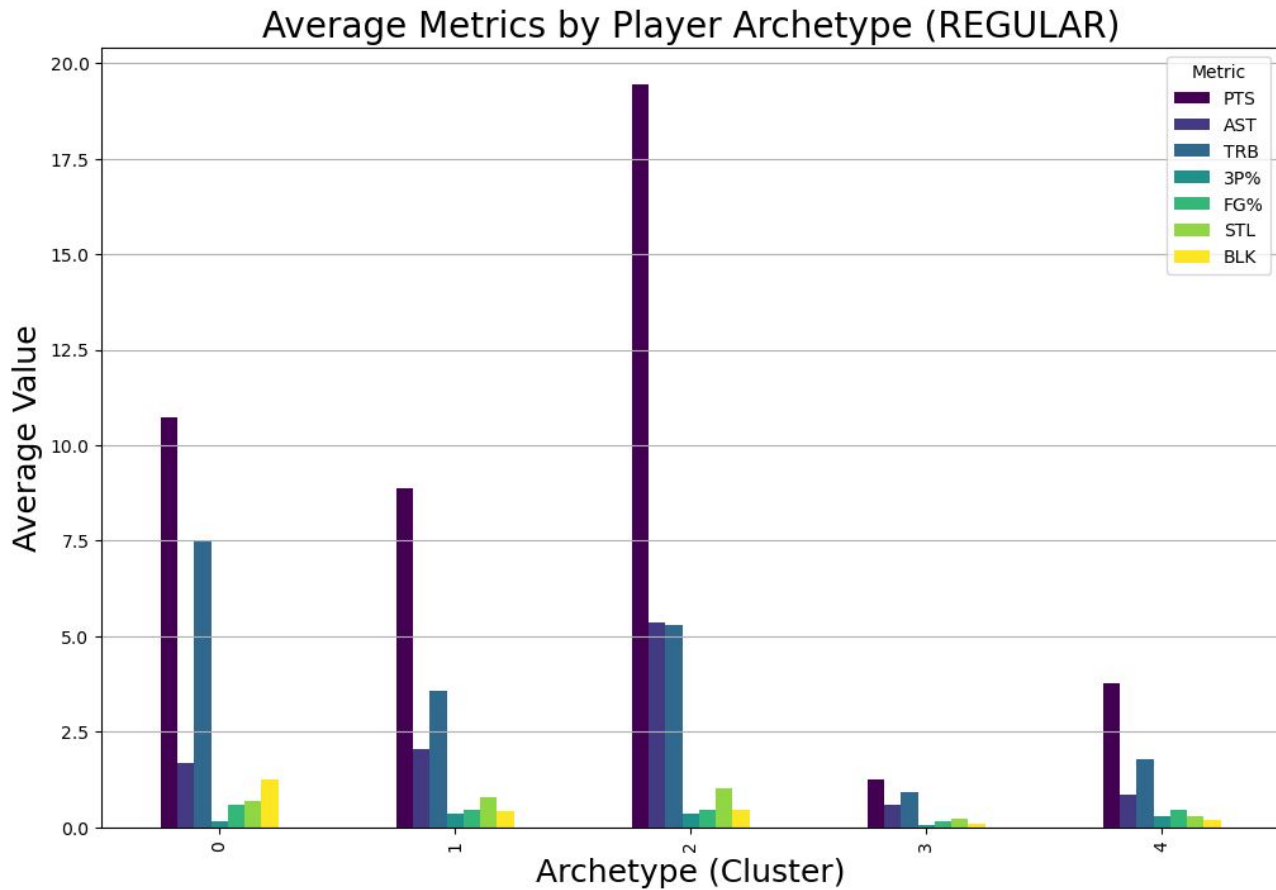


Cluster Evaluation

Refined Player Archetypes in 2D PCA Space (REGULAR)



Metrics By Player Archetype



Player Archetypes

“The Stars” – Highly efficient, play-making celebrities

“The Shields” – Defensive specialists/ball hawks

“The Snipers” – Excel at draining the long ball

“The Jacks” – Jacks of all trades

“The Bench” – Limited, low-impact role players

Feature Engineering

PPM = Points per game (PTS) / Minutes played per game (MP)

FG Miss = Field goals attempted (FGA) – Field goals made (FG)

FT Miss = Free throws attempted (FTA) – Free throws made (FT)

Effic = A measure of a player's on-court efficiency

Recov = Steals + Blocks + Total Rebounds – Turnovers

VPS = The player's Value Point System

Machine Learning

- Class Imbalances Handled Using SMOTE
- Dummy Classifier As Baseline Model
- Metrics Used:
 - Precision
 - Recall
 - F1 Score
 - ROC–AUC

Synthetic Minority Oversampling Technique

	3P%	BLK	STL
Real Star #1	0.358	0.9	1.4
Real Star #2	0.368	0.5	1.8
Synthetic Star	0.363	0.7	1.6

Baseline vs Final Model (JACK)

Scores On Validation Set:

	F1 Score	ROC–AUC
Dummy Classifier	0.17	0.52
CatBoost Classifier	0.93	0.96

Conclusion

Recall that our goal was to build a team where...

- There are 3 players of each of the 5 archetypes
- The team's total salary is at most \$140 million
- The players are as young as possible

The Final Draft

BENCHES: S. Cissoko, R. Rupert, J. Walsh

JACKS: B. Coulibaly, B. Miller, V. Wembanyama

SHIELDS: J. Champagnie, J. Giddey, A. Pokusevski

SNIPERS: L. Miller, R. Rollins, C. Wallace

STARS: S. Barnes, A. Edwards, J. Williams

Total Team Salary = \$83,262,908