Scouting Anyone: Probabilistic Player Archetypes for Any League

Paper Track: Basketball
Paper ID: 57

1.   Introduction

Classifying player roles is a central challenge in basketball analytics. While the sport is often described in terms of traditional positions (guard, forward, center), modern basketball has evolved into a continuum of playing styles shaped by spacing, pace, and tactical diversity. Players increasingly occupy hybrid roles, and their responsibilities change across teams, coaches, and developmental stages. As a result, role classification requires models that capture nuance, flexibility, and evolution over time, rather than rigid or subjective labels.

Elite organizations in leagues such as the NBA have adopted advanced scouting infrastructure and player-tracking technologies to support role identification and recruitment. However, most professional and semi-professional teams globally operate with limited budgets, limited staff, and minimal proprietary data. These teams often rely on qualitative scouting reports or simple statistical comparisons, which can be inconsistent across contexts and fail to identify players who could excel in new roles.

Another problem is that existing clustering approaches in basketball analytics, such as k-means or hierarchical clustering, typically assign each player to a single group. This is problematic: a stretch big who posts up occasionally and plays pick-and-roll defense does not strictly belong to one role. Such methods also gravitate toward "average" player types, producing centroids that do not align with real basketball archetypes. They cluster around the middle, while basketball roles are more meaningfully described at the extremes.

In this work, we apply Archetypal Analysis (AA) and Archetypoid Analysis (ADA) to classify player roles in a way that (a) reflects real extremes of playing style, (b) allows each player to belong probabilistically to multiple archetypes, and (c) can be built using data that is easily accessible. AA identifies extreme statistical profiles, or archetypes, while ADA ensures these archetypes correspond to actual players, improving interpretability for coaches and scouting staff. Each player-season-team profile is then modeled as a convex combination of archetypes, producing a role distribution rather than a single assignment. We showcase 2 examples, one using AA on European basketball data and the other using ADA on NBA data.

Our results demonstrate that this approach consistently produces coherent, interpretable role archetypes across both European and NBA datasets. Moreover, the framework is scalable, it works with simple statistical data, yet improves further as richer datasets (such as play-type distributions or shooting location profiles) are added. The models can be effectively applied to any league or level of competition, provided they are trained on data from leagues of comparable quality. The methodology therefore provides a practical and cost-effective decision-support tool for teams across leagues and resource environments.

1

Beyond methodological contributions, the probabilistic archetype framework also generates direct, operational value for basketball organizations. By quantifying stylistic similarity and lineup context, the model supports several practical applications: diagnosing role mismatches within lineups or rotations, identifying stylistically comparable players for scouting across leagues, and informing balanced roster construction. These applications connect the analytical foundations of the method to real decision-making processes in coaching, recruitment, and front-office strategy.

To demonstrate this in practice, the paper includes a detailed case study using 2024-25 EuroLeague data, where the framework is used to diagnose a structural mismatch involving a high-usage perimeter shooter and his surrounding lineup ecosystem. Additional examples illustrate how archetype analysis can guide roster design and support cross-league player identification. Together, these applications show how archetype-based modeling can serve as an accessible, interpretable, and actionable decision-support tool for teams across varied competitive and resource environments.

2.   Related Work & Background

Classifying basketball players into meaningful roles has long been of interest to both researchers and practitioners. Traditional methods in basketball analytics have relied heavily on manual scouting taxonomies, where players are labeled using position-based heuristics (e.g., guard, forward, center) or role descriptions crafted by coaching staff and scouts. As the modern game evolved toward pace-and-space offense and switching defenses, these positional labels became less adequate, motivating research into role-based rather than position-based classification. While these taxonomies are useful, they tend to be subjective, vary significantly across leagues and organizations, and struggle to capture the evolving and multi-dimensional nature of modern basketball.

To formalize role identification, analysts have commonly used clustering techniques such as k-means, hierarchical clustering, and Gaussian Mixture Models (GMMs) to group players based on statistical similarity. These approaches have been applied to box score statistics, lineup impact metrics, and tracking data. However, each comes with limitations in representing how roles function in practice. K-means, for example, assigns each player to exactly one cluster, implying a single, fixed role, an assumption that rarely reflects reality. GMM-based approaches, including prior work such as the work from Kalman & Bosch (1), do allow for probabilistic or "soft" role memberships, but the resulting assignments typically remain highly peaked (e.g., a player is ~90% associated with one cluster and only marginally with others). This is problematic because player roles are inherently fluid: responsibilities shift depending on teammates, coaching schemes, lineup context, and league style. A player may be a secondary creator in one environment and primarily a floor spacer in another. Traditional clustering methods tend to compress these nuanced role overlaps into overly rigid labels.

Archetypal Analysis (AA), introduced by Cutler and Breiman (1994) (2), offers a different approach. Instead of finding average or central clusters, AA identifies extreme points in the dataset, theoretical endpoints of playing styles, and represents each observation as a convex combination of these extremes. In basketball, this aligns with how coaches and analysts often describe players relative to prototypes (e.g., "a smaller Jokic-type facilitator," "a more mobile rim protector"). This convex

combination property enables probabilistic role representation, acknowledging that players rarely fit neatly into single categories.

(a) Data set in x,y plain

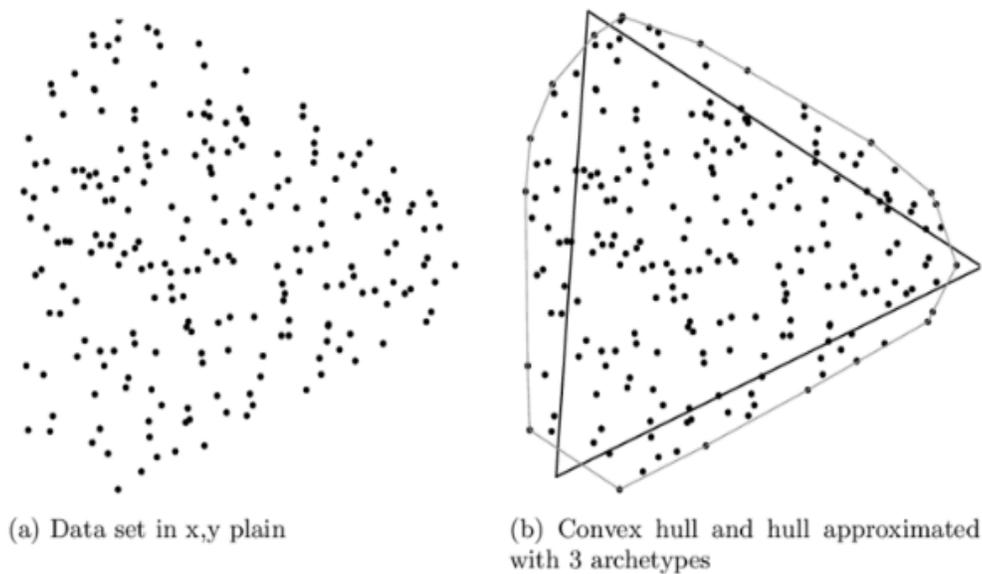(b) Convex hull and hull approximated with 3 archetypes

Figure 1: How Archetypal analysis approximates the convex hull of the dataset (3)

Archetypal analysis has been applied extensively in psychology (personality profiles), sociology (group identity modeling), economics (consumer behavior segmentation), and environmental science (extreme climate pattern identification). In these areas, AA is valued because it reveals interpretable prototype behaviors while allowing individuals to exist between extremes. However, its use in sports analytics has been minimal. Basketball is a domain where stylistic blending is the norm, making AA particularly well-suited since roles are fluid, lineup constructions are flexible, and the same player may have different responsibilities across teams, leagues, or seasons.

A related development is Archetypoid Analysis (ADA) (Vinué, Epifanio & Alemany 2015) (4), which constrains archetypes to be actual observed players instead of synthetic combinations. ADA enhances interpretability, particularly for practitioners who benefit from being able to say, for example, "Player X is stylistically 40% similar to Player Y and 35% to Player Z", rather than referencing abstract archetype centroids. This distinction is crucial for scouting departments and front offices, where interpretability is often valued as highly as statistical accuracy.

Recent work in basketball analytics has explored the use of tracking data, play-type frequencies, and spatial shot profiles to refine player style classification, particularly in the NBA where advanced data infrastructure is widely available. However, such data is not accessible to most international leagues. This creates a resource accessibility gap, where elite teams benefit from rich analytics, while the majority of professional teams must rely on limited box score statistics. The challenges associated with cross-league comparisons further complicate scouting markets and player movement.

The present research builds on this foundation by integrating AA and ADA into a unified, probabilistic framework for player role classification that works across leagues regardless of data richness. By structuring observations at the player-season-team level and drawing upon predictors that are consistently available worldwide, this methodology enables scalable, reproducible, and interpretable role profiles. Furthermore, by allowing players to hold probabilities across multiple archetypes, the approach captures stylistic fluidity and evolution over time, a central limitation of previous hard-clustering approaches.

3. Data & Preprocessing

3.1 Observation Structure: Player–Season–Team Units
Because a player's role can change depending on roster context, coaching philosophy, age, or league environment, our unit of analysis is a player–season–team combination, rather than a player in aggregate. For example, the same player may appear as a high-usage primary scorer in one season and as a spot-up complementary shooter in another. This structure allows the model to capture role evolution, making the archetypes reflect how players actually functioned in a given context, rather than how they are perceived in general.

3.2 European Dataset (2014-15 season – 2024-25 season)
We compiled box-score and advanced game-level statistics from seven major European professional leagues using publicly available data from RealGM.com. These leagues included the Spanish (ACB), French (LNB), Italian (LBA), German (BBL), Turkish (BSL), Greek (GBL) and Adriatic League (ABA) as they represent the highest competitive tier within their respective countries and feature comparable levels of play. This ensures consistency in player role distributions and statistical comparability across contexts. We deliberately excluded pan-European competitions such as the EuroLeague, EuroCup, and FIBA Basketball Champions League from training the model, as their substantially higher level of competitiveness and talent concentration could bias the archetype formation process. After aggregating to season averages and normalizing per-possession where appropriate, the following features were included:

List of features included in the European Dataset

Category	Variables
Physical	Height
Rebounding	Offensive Reb%, Defensive Reb%
Playmaking	Assist%, Turnover%
Defensive	Steal%, Block%
Usage & Efficiency	Usage%, Free Throw Rate, FT%, 2P%, 3P%
Scoring Profile	Points per 100 poss., FGA per 100 poss., Share of 3PA

These features were selected to capture how players impact possessions (creation vs. finishing), where they operate on the floor, and how efficiently they convert opportunities.

3.3 NBA Dataset (2017-18 season – 2024-25 season)

For the NBA dataset, we extended the feature space to include shot distribution and play-type behavior, using data from NBA.com and Basketball-Reference.com.
In addition to the features included in the European model, the following features were also included:

List of additional features included in the NBA Dataset

Category	Variables
Shot Distribution	%FGA at 0-3 ft, 3-10 ft, 10-16 ft, 16 ft-3P, above-the-break 3s, corner 3s, average shot distance
Assisted vs Unassisted Shots	%AST-2P, %AST-3P
Play Types	% of possessions via pick-and-roll (handler/roller), isolation, handoffs, post-ups, cuts, off-screen, putbacks, spot-ups

These features provide insight into role execution, distinguishing, for instance, a pick-and-roll creator from a movement shooter or interior finisher.

3.4 Normalization and Scaling
Because the variables in our dataset span different numerical scales (e.g., percentages, counts, centimeters), we apply z-score normalization prior to modeling. This step ensures that no single feature disproportionately influences the analysis solely due to its magnitude. Standardizing places all variables on a comparable scale, allowing the archetypes to meaningfully reflect underlying playing styles rather than measurement differences.

3.5 Filtering for Stability
We apply a minimum playing-time threshold of 300 minutes played per player-season-team in both the European and NBA datasets to ensure statistical stability. This reduces noise and prevents role patterns from being driven by players with very limited sample sizes, where performance tends to exhibit high variance. By filtering out short-minute or injury-shortened seasons, we ensure that the archetype representations reflect sustained playing tendencies rather than small-sample fluctuations.

4.   Methodology

The objective of this study is to identify interpretable, probabilistic player role profiles derived directly from observed performance data. To achieve this, we applied Archetypal Analysis (AA) and Archetypoid Analysis (ADA) to our standardized player–season–team datasets. Both approaches model players not as members of a single cluster, but as convex combinations of stylistic extremes, enabling more realistic representation of hybrid and evolving playing styles.

4.1 Archetypal Analysis (AA)
Archetypal Analysis seeks to represent each observation x_i in the dataset as a weighted mixture of a small set of archetypes $Z = \{z_1, z_2, ..., z_K\}$, where each archetype itself is a convex combination of

data points. Formally, AA starts with a dataset X of n observations (players) and m variables (performance metrics). The method searches for k archetypes, represented by the matrix Z, that best approximate all data points. Each archetype is a convex combination of real players, and each player is in turn a convex combination of these archetypes. Mathematically:

$$||X - \alpha Z||^2, \text{ where } Z = \beta X$$

subject to:

$$\alpha_{ij} \geq 0, \sum_{j} a_{ij} = 1 \text{ and } \beta_{ji} \geq 0, \sum_{i} b_{ji} = 1$$

where:
- α_{ij} represents the mixing weights (how much player i belongs to archetype j)
- β_{ji} the coefficients used to define each archetype as a combination of actual data points
- The constraints ensure that all players and archetypes are weighted averages (no negative weights, total weight = 1).

The archetypes represent extreme, interpretable stylistic endpoints rather than centroids. Player memberships α_{ij} form probabilities (e.g., 0.60 3P Specialist, 0.25 Playmaker, 0.15 Defensive Wing) and every player can express a blend of roles, reflecting real-world basketball dynamics. This differs fundamentally from clustering methods like k-means, where players are forced into a single category, even when their style is hybrid.

4.2 Archetypoid Analysis (ADA)
Archetypoid Analysis modifies AA by constraining archetypes to be actual players rather than theoretical convex combinations. This enhances interpretability and practical usefulness. Mathematically, we still have the same minimizing function:

$$||X - \alpha Z||^2, \text{ where } Z = \beta X$$

but the constraints are slightly different:

$$\alpha_{ij} \geq 0, \sum_{j} a_{ij} = 1 \text{ and } \beta_{ji} \in \{0, 1\}, \sum_{i} b_{ji} = 1$$

The key difference from AA lies in the binary constraint on β:
- Each archetypoid Z_K is a *single actual data point* (a real player), because $\beta_{ji} = 1$ only for one observation i, and 0 otherwise
- Each player X_i is still a convex combination of these archetypoids via α_{ij}

ADA is particularly effective for scouting and communication, as archetypes can be described with statements such as "This player is 40 % stylistically similar to Player Y and 35% to Player Z."

4.3 Determining the Number of Archetypes (K)

Using the standard elbow method on reconstruction error (RSS) provided an initial guideline for selecting the number of archetypes. However, in both datasets the elbow appeared very early, around four archetypes, which is clearly insufficient to capture the diversity of modern basketball roles. While four archetypes minimize RSS efficiently, they collapse multiple distinct functional player types into overly broad categories (e.g., combining creators, shooters, and hybrid forwards), limiting interpretability and practical value.
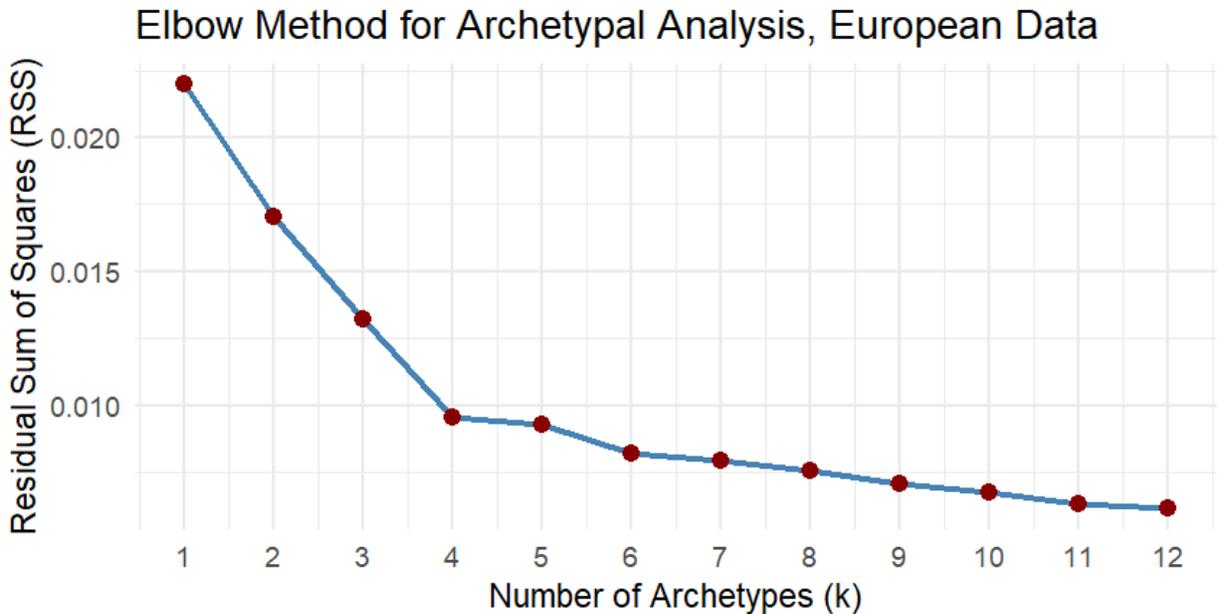

Figure 2: Scree Plot, European Dataset, Archetypal Analysis (AA)

To address this, we introduced an additional diagnostic: intra-archetype variance. Instead of focusing on global reconstruction error, this metric evaluates how similar the players assigned to each archetype are to one another. Lower intra-archetype variance indicates that archetypes capture more coherent, internally consistent roles. By examining how this variance changes as the number of archetypes increases, we can detect when additional archetypes meaningfully refine role structure, and when they simply overfit noise.

Applying intra-archetype variance analysis to the European dataset and the AA methodology revealed a clear optimum at 8 archetypes. Variance steadily decreased as we increased the number of archetypes from 4 to 8, reflecting that each additional archetype helped isolate more precise on-court behaviors. Importantly, after k = 8, variance increased considerably, and further archetypes mostly subdivided already coherent roles rather than uncovering new, meaningful distinctions. In other words, eight archetypes represented the point where the model captured the essential diversity of player types without fragmenting them into unnecessarily fine-grained categories. This aligns with qualitative evaluation: each of the eight roles corresponded to recognizable, interpretable basketball archetypes with minimal redundancy.
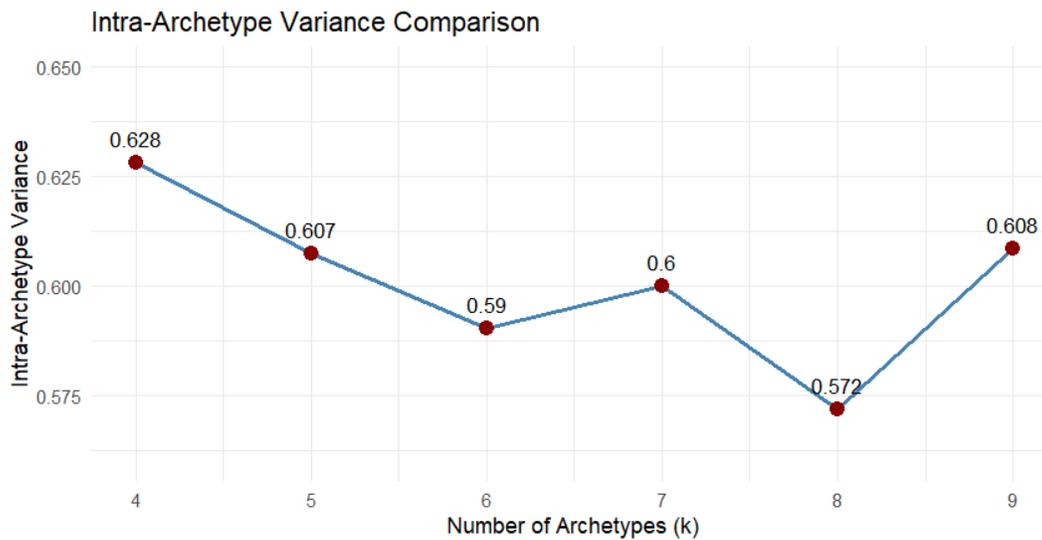
Figure 3: Intra-archetype Variance per number of Archetypes, European Dataset, Archetypal Analysis (AA)

For the NBA dataset and the ADA methodology, intra-archetypoid variance followed a similar pattern, reaching its minimum at 9 archetypoids. However, the improvement from 8 to 9 archetypoids was marginal, the reduction in variance was noticeably smaller than in earlier steps, decreasing from 0.538 to 0.525, and the ninth archetypoid often appeared as a subtle variation of an existing one rather than a distinct, stable role thus leading us to adhere to 8 archetypoids. We refrained from testing much larger numbers of archetypes or archetypoids, as both basketball domain knowledge and prior literature suggest that the ecosystem of meaningful player roles is not dramatically larger. Pushing the model further would produce increasingly niche, fragmented groups with limited interpretability.
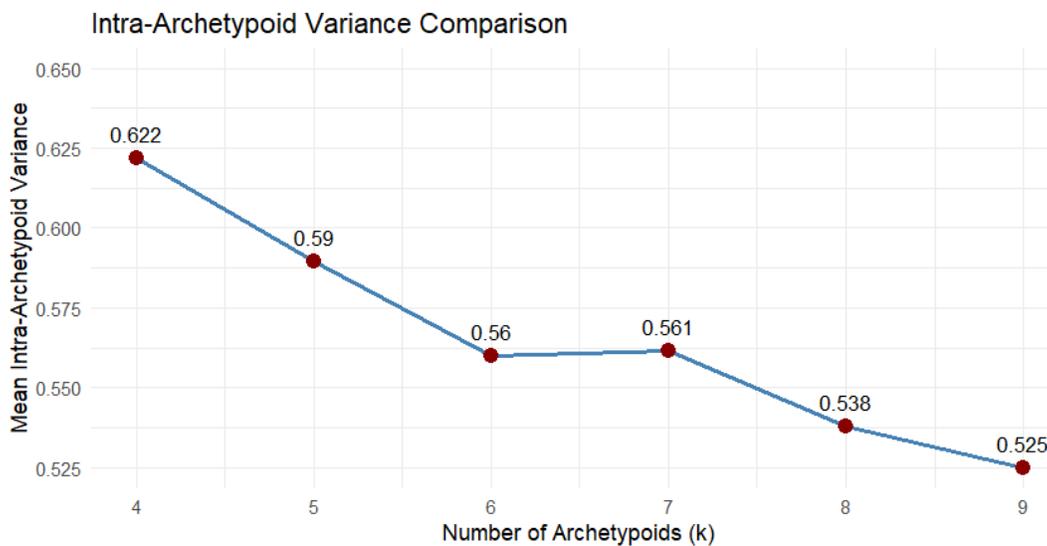

Figure 4: Intra-archetypoid Variance per number of Archetypoids, NBA Dataset, Archetypoid Analysis (ADA)

5.   Results

5.1 European Archetypes (AA Model)

Since the labels of the 8 clusters do not correspond to predefined roles, we carried out a detailed exploratory analysis to interpret the player types represented by each group. This included reviewing descriptive statistics, examining mean z-scores for players most strongly associated with each archetype, and using boxplots to identify which variables separated one group from another. The table below highlights the eight archetypes derived from the European dataset, summarizing their characteristics and the features that most clearly distinguish them.

Archetype	Explanation	Distinguishing Features	Top Players
High Usage Guard	Primary creator and scorer	High Usage%, High Points per 100, High AST%	N. Hilliard (GER 2020-21) D. Russell (GER 2022-23) M. Huertas (ESP 2021-22)
3PT Specialist	Off-ball shooter with limited on-ball creation	High 3FGA%, High 3P%, Low Free Throw Rate	M. Mahmutoglu (TUR 2022-23) B. Heslip (TUR 2018-19) A. Corbacho (ESP 2018-19)
Traditional Center	Interior finisher + rim protection	Taller, High Reb%, High BLK%, High FTr, Low 3PA%	E. Arar (TUR 2023-24) S. Spencer (ITA 2023-24) D. Naymick (GRC 2014-15)
Backup Big	Low-usage interior role player	Similar to Traditional Center but with lower efficiency and usage	O. Mackeldanz (DEU 2018-19) S. Erden (TUR 2019-20) M. Lessort (Adriatic 2022-23)
Defensive Specialist	Disruptive defender with low offensive burden	High STL% and/or BLK%, Low Usage%	B. Cerella (ITA 2017-18) K. Tadda (DEU 2014-15) K. Tadda (DEU 2020-21)
High Usage Forward	Multi-level scoring forward	Taller than the average, High Usage%, High Points per 100, High FGA per 100, Low 3FGA%	F. Petrusev (Adriatic 2020-21) J. Tillman (GRC 2023-24) B. Dubljevic (ESP 2015-16)
Traditional Playmaker	Pass-first facilitator	Shorter, High AST%, High TOV%, Low scoring production	D. Karamfilov (Adriatic 2014-15) A. Kerckhof (FRA 2017-18) L. Dercole (ITA 2016-17)

| Role Guard | Complementary backcourt piece | Shorter, Average in most areas, slightly elevated usage | D. Ozdemiroglu (TUR 2019-20)
O. Williams (DEU 2022-23)
J. Morgan (FRA 2023-24) |

Leagues References: GER: German BBL; TUR: Turkish BSL; ESP: Spanish ACB; ITA: Italian Serie A; Adriatic: Adriatic League; FRA: French LNB Elite

The figure below illustrates the average z-scores across all variables for two example archetypes, Traditional Centers and Traditional Playmakers.
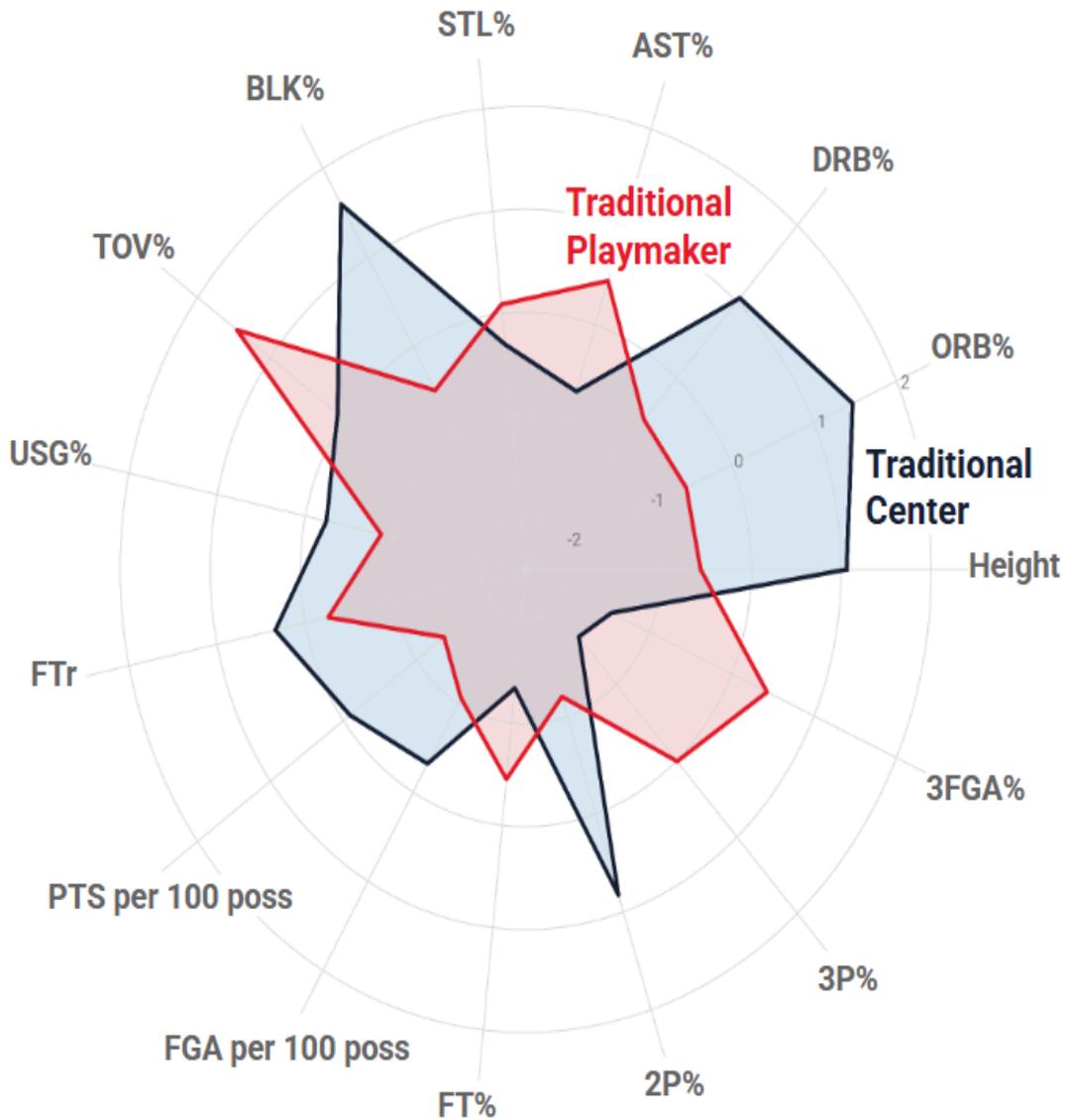

Figure 5: Traditional Center vs Traditional Playmaker Normalized performance - Average Z-scores per variable

Below we have the results for Dylan Osetkowski from the 2023-24 season with Unicaja Malaga as an example.

Figure 6: Dylan Osetkowski's Archetype Analysis results - 2023-24 Liga Endesa season, Unicaja Malaga

As discussed above, instead of assigning a single label, each player-season receives a vector of membership weights. For instance, Dylan Osetkowski's playing style in 2023-24 can be described as:
- 57.0% High Usage Forward
- 23.2% Role Guard
- 15.2% 3PT Specialist and
- 4.5% Defensive Specialist

5.2 NBA Archetypes (ADA Model)
Similarly, for our Archetypoid Analysis model, trained on NBA data with an enriched dataset, we performed the same EDA and we identified the following 8 archetypoids:

Archetypoid	Explanation	Distinguishing Features	Top Players
Inside Scoring Big	Interior finisher and offensive rebounder	Tallest, Highest %Post-Up, High %Roll Man and Putback, High USG% and scoring production	N. Vucevic (2020-21) L. Aldridge (2017-18) N. Vucevic (2020-21)
Shooting Specialist	Movement or spot-up shooter	High Off-Screen & Spot-up, Low self-creation, Low 2FGA%	T. Daniels (2017-18) W. Ellington (2019-20) W. Ellington (2017-18)
Combo Guard	Scoring guard with creation ability	Shorter, High PnR Handler% & Isolation%, moderate AST%	T. Burke (2021-22) J. Crawford (2017-18) T. Burke (2018-19)
Offensive Engine	Primary initiator and focal scorer	Very high Usage%, High scoring production, Low % of assisted shots, isolation % PnR-heavy	J. Harden (2018-19) L. Doncic (2022-23) J. Harden (2019-20)
Traditional Playmaker	Pass-first facilitator	High AST%, Higher TOV%, Low shot volume, High PnR Handler%	E. Payton (2024-25) I. Smith (2023-24) R. Rondo (2020-21)
3&D Wing	Perimeter defender & floor spacer	Average Size, High 3PA%, Moderate, STL% and BLK%, Minimal creation or Rim presence	D. Wade (2024-25) D. Wade (2022-23) K. Edwards (2022-23)
Rim Protector / Roll Man	Interior defender + PnR finisher	Elite BLK% & OREB%, High Roll Man and Putback frequency	M. Robinson (2022-23) M. Robinson (2021-22) M. Robinson (2023-24)
Mobile Big	Stretch-capable or perimeter-mobile big	Slightly Shorter and other Bigs, High Corner 3FGA%, High % of Assisted FGA, High% on Cuts	J. Poeltl (2017-18) O. Okongwu (2022-23) J. Allen (2017-18)

Below we have the results for Giannis Antetokounmpo from the 2024-25 season with the Milwaukee Bucks.

Giannis Antetokounmpo
NBA 2024-25

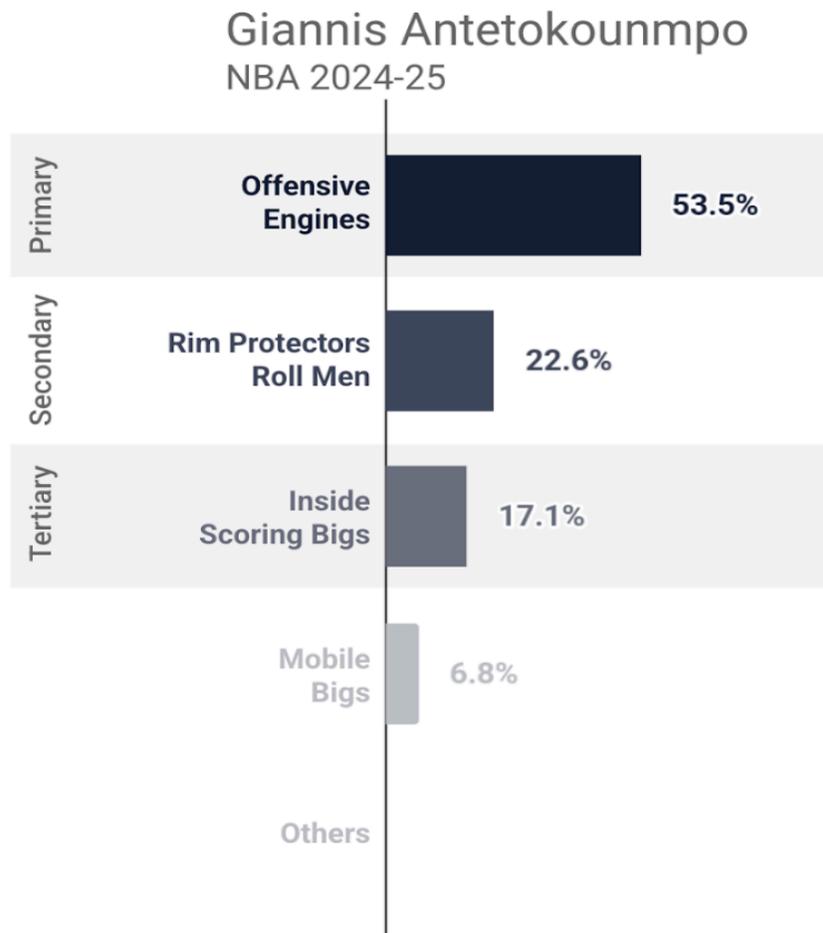|	Archetype	Percentage
Primary	Offensive Engines	53.5%
Secondary	Rim Protectors Roll Men	22.6%
Tertiary	Inside Scoring Bigs	17.1%
	Mobile Bigs	6.8%
	Others	

Figure 7: Giannis Antetokounmpo's Archetypoid Analysis Results - 2024-25 NBA Regular Season, Milwaukee Bucks

Giannis Antetokounmpo's playing style in 2024-25 can be described as:
- 53.5% similar to 2019 James Harden and 2023 Luka Doncic (Offensive Engine)
- 22.6% similar to 2022, 2023 Mitchel Robinson and 2022 DeAndre Jordan (Rim Protector/Roll Man)
- 17.1% similar to 2021 Nikola Vucevic (Inside Scoring Big) and
- 6.8% similar to 2018 Jakob Poeltl and Jarrett Allen and 2023 Onyeka Okongwu (Mobile Big)

6.   Application for Teams

The probabilistic archetype framework developed in this research enables several practical applications that address real operational challenges faced by basketball organizations. In this section, we demonstrate three key use cases:

1. Diagnostic analysis: Identifying player combination mismatches in existing rosters
2. League-wide or cross-league scouting: Finding stylistically similar players within the league/competition or across different leagues/competitions

3. Roster construction: Balancing archetype distributions for tactical coherence

Each application leverages the core methodological advantages established in Section 4: probabilistic role representation, interpretability through real player archetypes (ADA) or extremal profiles (AA), and scalability across data richness levels.

We present a detailed case study of the first application (player combination mismatch detection) using 2024-25 EuroLeague data to demonstrate how archetype analysis can diagnose performance issues and guide tactical adjustments. This is followed by brief illustrations of roster construction applications.

6.1 Case Study: Diagnosing Player Combination Mismatch in High-Usage Lineups

6.1.1 Context and Methodology

For this application, we retrained our Archetypal Analysis model on EuroLeague-specific data from the 2024-25 season (N=249 player-season observations, minimum 300 minutes played through Regular Season). While our European dataset (Section 3.2) was trained on domestic leagues, EuroLeague represents a higher competitive tier featuring many of the same players and tactical systems. The resulting eight archetypes closely mirror those identified in the NBA (Section 5.2)

EuroLeague Archetype Definitions: The retrained model identified the following eight archetypes, which we use throughout this case study:

Archetype Name	Explanation	Distinguishing Features
Shooter Specialist	Elite perimeter shooter, off-ball oriented	High 3FGA%, high spot-up frequency, minimal self-creation
3&D Wing	Two-way perimeter player	Moderate 3PA%, high STL%, low creation rate, defensive versatility
Lead Guard	Pass-first facilitator	Very high AST%, high TOV%, low scoring volume, high pick-and-roll handler usage
Mid-Range Specialist	Interior slasher with midrange game	High 2P% from 10-16ft, low 3PA%, moderate USG%, pull-up oriented
Traditional Center	Interior finisher and rim protector	Tallest, high BLK%, high OREB%, low 3PA%, post-up and putback frequency
High Volume Scorer	Multi-level scoring threat	High USG%, high points per 100, balanced shooting from multiple areas
Stretch 4	Stretch big with perimeter shooting	High corner 3PA%, high pick-and-pop frequency, moderate height, floor spacing
Rolling Big	Vertical spacer and dunker	High roll-man%, high putback%, high FG% at rim, lob threat, minimal perimeter shooting

6.1.2 Player Profile: Markus Howard's Extreme Archetype Concentration

Baskonia is a high-caliber basketball team in Spain which competes in the Spanish National League (Liga ACB) as well as in the EuroLeague. Baskonia features a high-usage perimeter scorer whose stylistic identity is exceptionally well-defined. Markus Howard, a former NCAA scoring leader at Marquette (leading scorer with 27.8 PPG in 2019-20) and former Denver Nuggets guard, was classified by our EuroLeague model as 72% Shooter Specialist, one of the highest single-archetype concentrations observed in the league.
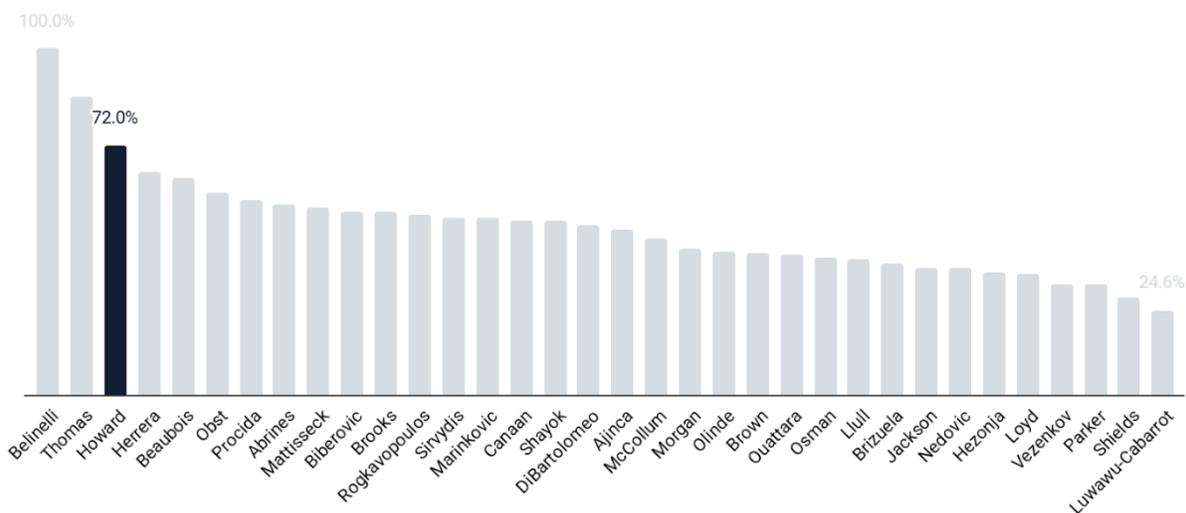

Figure 8: Shooter Specialist Archetype Probability Distribution in 2024-25 EuroLeague

Statistically, Howard embodies the modern high-volume shooter profile:

- 12.1 points per game (89th percentile among 2024-25 EuroLeague players with ≥ 100 minutes)
- 7.4 three-point attempts per game (99th percentile)
- 2.5 made threes per game (97th percentile)
- 33.9% from three (44th percentile overall; 26th-30th percentile among high-volume shooters with ≥ 3 3PA/game)

This combination of extreme volume with league-average efficiency is characteristic of shooters whose primary value derives from gravity (defensive attention commanded) rather than raw shooting percentages. Players with such extreme perimeter profiles typically require specific structural environments to maximize efficiency: secondary ball-handlers capable of generating advantages, stretch forwards or mobile bigs who preserve spacing, and wings who provide both shooting and defensive coverage.

When these complementary roles are absent or misaligned, teams often struggle to convert the shooter's gravity into collective efficiency, leading to diminished offensive performance and negative lineup outcomes.

6.1.3 Problem Identification: Performance Split Signals Structural Issue

Despite Howard's individual offensive contributions, Baskonia's team performance exhibits a stark contrast based on his court presence:

- **Net Rating[1] with Howard on court**: –7.0
- **Net Rating with Howard off court**: +3.0

This represents a 10-point swing per 100 possessions, a substantial margin at the EuroLeague level. Importantly, this does not suggest Howard is a negative-impact player in isolation. Rather, it indicates a deeper structural issue: the lineups surrounding him are not aligned with the requirements of his archetype.

6.1.4 Ecosystem Analysis: Who Plays Alongside Howard?

To understand why Baskonia underperforms with Howard on the floor, we examine the archetype distribution of his most frequent teammates at two levels: macro (archetype exposure patterns) and micro (individual player profiles).

6.1.4.1 Macro View: Archetype Distribution in Howard's Minutes

We calculated the percentage of Howard's total court time (715 minutes in 2024-25 EuroLeague Regular Season) shared with each of the eight EuroLeague archetypes. The distribution reveals a clear structural imbalance:

High-Exposure Archetypes (collectively 57.3% of shared time):
- **Traditional Center**: 25.8% of Howard's minutes
- **Mid-Range Specialist**: 16.9%
- **High Volume Scorer**: 14.6%

Low-Exposure Archetypes (collectively 21.7% of shared time):
- **3&D Wing**: 9.5%
- **Stretch 4**: 7.2%
- **Rolling Big**: 3.0%

---

[1] Net Rating is calculated as the difference between Points scored per 100 possessions minus Points conceded per 100 possessions.

Figure 9: Markus Howard's Archetype Exposure Patterns Across 2024-25 EuroLeague Season

This pattern is problematic for a 72% Shooter Specialist. More than half of Howard's shared playing time comes with archetypes that compress offensive spacing (Traditional Centers operate almost exclusively in the paint, Mid-Range Specialists pull defenders inside the arc) or provide limited perimeter gravity (Lead Guards with minimal shooting threat). In contrast, the archetypes that typically amplify high-volume shooters, 3&D wings, stretch bigs, and fellow shooters, represent less than 30% of his ecosystem.

This macro-level view suggests Howard consistently operates in lineups that do not match the structural requirements of his role, placing him in environments with multiple non-shooters that diminish the value of his gravity and increase defensive pressure on his actions.

6.1.4.2 Micro View: Howard's Three Primary Teammates

Across the season, Howard played 91.5% of his on-court minutes with at least one of the three teammates profiled below.

Chima Moneke (470 shared minutes, 65% of Howard's time):
- 43.4% Mid-Range Specialist
- 25.7% High Volume Scorer

Donta Hall (443 shared minutes, 62% of Howard's time):
- 82.4% Traditional Center
- 10.0% Stretch 4

Trent Forrest (403 shared minutes, 56% of Howard's time)
- 45.7% Lead Guard
- 27.0% High Volume Scorer

Chima Moneke

Mid Range Specialist		43.4%
High Volume Scorer		25.7%
Traditional Center		15.2%
Stretch 4		7.4%
Others		8.3%

Donta Hall

Traditional Center		82.4%
Stretch 4		10.0%
High Volume Scorer		7.3%
Rolling Big		0.2%
Others		0.1%

Trent Forrest

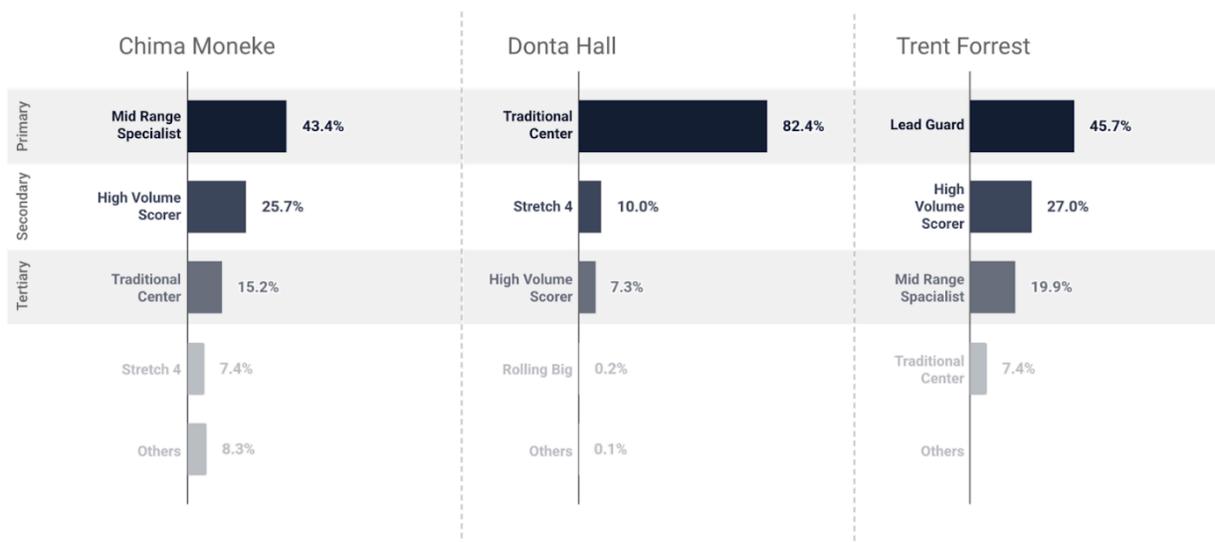|		
Lead Guard		45.7%
High Volume Scorer		27.0%
Mid Range Spacialist		19.9%
Traditional Center		7.4%
Others		

Figure 10: Role Composition of Howard's Most Frequent On-Court Teammates in the EuroLeague (2024-25)

While each player can be individually productive, their stylistic overlap creates the least compatible offensive environment for a high-volume Shooter Specialist. This archetype-level analysis reveals why traditional scouting, which might label Forrest as a "Point Guard" and Hall as a "Center", fails to capture the nuanced spacing incompatibility that affects lineup performance.

6.1.5 Lineup Performance: Quantifying the Mismatch

Using possession-adjusted net ratings across all 2024-25 EuroLeague Regular Season games, we analyzed Howard's most common lineup combinations with his three highest-minute teammates: Forrest, Hall, and Moneke. The results reveal a striking pattern: every major pairing yields a negative net rating.

Two Players Combinations

Combination	Minutes	% Minutes w/ Howard	ORtg	DRtg	NRtg
Howard + Moneke	470.0	65.8%	110.3	124.7	**-14.4**
Howard + Hall	443.0	62.0%	111.3	123.1	**-11.8**
Howard + Forrest	403.0	56.4%	115.9	124.3	**-8.4**

Three Players Combinations

Combination	Minutes	% Minutes w/ Howard	ORtg	DRtg	NRtg
Howard + Hall + Moneke	304.4	42.6%	105.9	125.0	**-19.1**

Howard + Forrest + Moneke	290.8	40.7%	109.6	124.9	**-15.3**
Howard + Forrest + Hall	257.3	36.0%	114.6	120.5	**-5.9**

This concentration indicates a highly stable rotation pattern, with nearly half of all team minutes (48%) involving at least one member of the Howard-Forrest-Hall-Moneke cohort.

The consistency across all configurations, regardless of which specific players are combined, suggests the issue transcends individual matchups. The archetype model provides the explanation: when Howard (72% Shooter Specialist) is surrounded by Forrest (45.7% Lead Guard, minimal shooting), Moneke (43.4% Mid-Range Specialist, interior-focused), and Hall (82.4% Traditional Center, paint-bound), the collective effect compresses spacing, reduces catch-and-shoot opportunities, and forces Howard into self-creation, outside his optimal role. These lineups systematically neutralize his primary value (gravity from volume shooting) without providing the structural support to convert that gravity into efficient offense.

6.1.6 League-Wide Comparison: Optimal Ecosystems for Shooter Specialists

To identify how Baskonia should construct lineups around Howard, we perform a league-wide analysis of players with similar archetypes and evaluate which environmental factors correlate with positive on-court performance.

6.1.6.1 Identifying High-Impact Comparables

We begin with the Top-20 guards and wings classified as Shooter Specialists (ranked by archetype probability $\geq 55\%$) who played ≥ 300 minutes in the 2024-25 EuroLeague Regular Season. We then filter this group to identify players who drive positive team impact, defined as meeting one or both criteria:

- Positive Net Rating when on court
- Higher team Net Rating on-court than off-court (positive on/off differential)

Among the top 20 Shooter Specialists, only eight players meet these performance criteria:

Player	Team	Shooter Specialist	Minutes	NRtg On	NRtg Off	NRtg Delta
Sebastian Herrera	Paris	64.6%	154	**+21.8**	**-2.4**	**+24.2**
Rodrigue Beaubois	Anadolu Efes	62.9%	623	**+6.4**	**+7.9**	**-1.5**
Alex Abrines	Barcelona	55.0%	563	**+9.9**	**+1.9**	**+8.0**
Tarik Biberovic	Fenerbahce	53.0%	718	**+8.5**	**-3.3**	**+11.8**
Nikos Rogkavopoulos	Baskonia	52.2%	618	**+6.8**	**-9.5**	**+16.4**
Vanja Marinkovic	Partizan	51.1%	481	**+6.7**	**-1.0**	**+7.7**

Isaiah Canaan	Crvena Zvezda	50.6%	711	**+3.2**	**+1.2**	**+2.0**
Errick McCollum	Fenerbahce	45.4%	238	**+3.5**	**+2.7**	**+0.8**

These eight players represent the benchmark for successful Shooter Specialist deployment in EuroLeague competition. Notably, Howard (72% Shooter Specialist, –7.0 on-court NRtg, –10.0 on/off differential) exhibits higher archetype concentration than any player in this successful cohort, yet dramatically underperforms them in team impact metrics.

6.1.6.2 Measuring Ecosystem Patterns

For each of the eight high-impact Shooter Specialists, we calculated the percentage of minutes shared with each archetype, weighted by shared court time. We then computed the unweighted mean across the cohort to produce a league baseline representing optimal Shooter Specialist environments.

The table below compares Howard's archetype exposure (playing time) against this baseline:

Archetype	League Baseline	Howard Actual	Difference (pp)	Interpretation
Traditional Center	12.9%	25.8%	**+12.9 ↑**	Howard plays far more minutes with interior-oriented bigs, reducing spacing
Mid-Range Specialist	14.7%	16.9%	**+2.2 ↑**	Slight overexposure to non-shooting scorers
High Volume Scorer	16.9%	14.6%	**-2.3 ↓**	Mild underexposure to secondary creators
Lead Guard	12.0%	13.3%	**+1.3 ↑**	Slightly increased exposure to non-shooting facilitators
Shooter Specialist	13.4%	9.8%	**-3.7 ↓**	Underexposure to spot-up shooters and spacing threats
3&D Wing	13.7%	9.5%	**-4.1 ↓**	Notable shortage of two-way perimeter defenders/spacers
Stretch 4	11.1%	7.2%	**-3.9 ↓**	Lack of stretch bigs limits pick-and-pop spacing
Rolling Big	5.4%	3.0%	**-2.4 ↓**	Underuse of vertical threats, reducing rim-pressure synergy

Key Findings:
1. Howard plays far more with spacing-negative archetypes:
- +12.9pp exposure to Traditional Centers (Hall's 82.4% archetype)

- +2.2pp exposure to Mid-Range Specialists (Moneke's 43.4% archetype)

These archetypes compress offensive spacing, place additional defenders near the basket, and reduce the effectiveness of Howard's perimeter gravity.

2. Howard plays far less with spacing-positive archetypes:
- -4.1pp less exposure to 3&D Wings (should be 13.7% of minutes, currently only 9.8%)
- -3.9pp less exposure to Stretch 4s (should be 11.1%, currently only 7.2%)
- -3.7pp less exposure to fellow Shooter Specialists (should be 13.4%, currently only 9.8%)

6.1.6.3 Validating Predictions with Howard's Actual Teammates

To confirm that the model's predictions translate into real outcomes, we tested three archetype pairings using Howard's existing roster and their actual on/off performance data:

Pairing	Primary Archetype	Howard + Player NRtg	Howard + Player Minutes	Only Howard NRtg	Only Player NRtg	Interpretation
Donta Hall	82.4% Traditional Center	-11.8	443	+0.7	+6.1	Negative interaction: both perform better apart, validating spacing compression prediction
Chima Moneke	43.4% Mid-Range Specialist	-14.4	470	+7.2	+7.2	Strong negative fit despite positive solo performance for both
Nikos Rogkavopoulos	52.2% Shooter Specialist	0.0	283	-11.6	+12.3	Howard improves +11.6 pts/100 with fellow shooter; best available partnership on roster
Sander Raieste	52.0% 3&D Wing	+22.5	19	-7.8	-25.6	+30.3 pt swing suggests strong synergy; limited sample requires validation over 150-200 minutes

Key Findings: The analysis confirms the model's expectations: Howard's performance declines sharply with Traditional Centers and Mid-Range Specialists, even though each player performs well independently. In contrast, lineups featuring shooting- and spacing-positive archetypes show strong improvements, with Shooter Specialists stabilizing Howard's impact and 3&D Wings generating the highest projected gains. The Raieste-Howard pairing, currently with only 19 minutes, should be expanded to validate the predicted synergy.

6.1.7 Roster Construction Gaps and Recruitment Priorities

Comparing Howard's archetype exposure patterns with the league baseline for successful Shooter Specialists reveals two critical roster deficiencies that limit Baskonia's ability to optimize their primary perimeter scorer.

Gap 1: No Rotation-Level Stretch 4
- **League average exposure**: 11.1% of Shooter Specialist minutes

- **Baskonia current exposure**: 7.2%
- **Deficit**: -3.9pp below optimal

Howard's successful peers (Beaubois, Abrines, Herrera, etc.) rely heavily on stretch bigs who can space the floor from the dunker spot, pop to the perimeter after screening, or provide corner three-point gravity. These players create vertical spacing that prevents defenders from collapsing on perimeter actions.

Baskonia's only player classified in this archetype is Ousmane Ndiaye (53% Stretch 4), who has played only 12 games with 115 total minutes in the 2024-25 season, insufficient for rotation-level impact. The lack of a consistent Stretch 4 forces Howard into lineups dominated by Traditional Centers (Hall: 82.4%) who compress spacing.

Recruitment Priority: This represents the highest-priority addition for optimizing Howard's ecosystem. Target profile:
- $\geq 50\%$ Pick-and-Pop Big archetype concentration
- Roll-man capability for pick-and-roll versatility
- Defensive mobility to switch on perimeter actions

Gap 2: Limited 3&D Wing Depth
- **League average exposure**: 13.7% of Shooter Specialist minutes
- **Baskonia current exposure**: 9.5%
- **Deficit**: -4.2pp below optimal

Successful Shooter Specialists spend approximately one-seventh of their court time alongside true 3&D wings who provide both perimeter shooting and defensive coverage. This archetype is critical because it:
1. Maintains spacing without defensive liabilities
2. Allows defensive switching schemes that protect the Shooter Specialist
3. Provides secondary ball-handling without compressing perimeter space

Baskonia's roster includes only Sander Raieste (52.0% 3&D Wing) who fits this profile, yet he plays sparingly (19 shared minutes with Howard, 2% of total) despite strong on/off data (+22.5 NRtg together).

Recruitment Priority:
- Increase Raieste's role and minutes (internal solution)
- Target external addition with clear 3&D profile:
  - o $\geq 40\%$ 3&D Wing archetype concentration
  - o $\geq 35\%$ three-point percentage on moderate volume
  - o Positive defensive metrics (STL%, on-ball defense rating)

7.   Conclusion

This research demonstrates that Archetypal Analysis and Archetypoid Analysis provide a scalable, interpretable, and probabilistic framework for describing basketball player roles.
Unlike traditional clustering models, which force players into single categories, this approach captures role blending and role evolution, reflecting how modern basketball is actually played.

By requiring only minimal data, the method is accessible to teams across the global basketball ecosystem, from elite organizations to resource-limited clubs.
At the same time, incorporating more advanced contextual features (such as shot distribution and play-type involvement) further enhances precision, demonstrating the scalability and extensibility of the framework.

The practical applications demonstrated in this work underscore the usefulness of archetype modeling for real basketball decision-making. The framework helps teams identify undervalued or context-dependent players across leagues, diagnose stylistic redundancies or mismatches within lineups, and guide roster construction toward more coherent tactical structures. Because the model is both reproducible and interpretable, particularly through ADA's use of real player examples, it integrates naturally into coaching, scouting, and front-office workflows.

In a sport defined by fluidity rather than fixed positions, this methodology provides a modern, practical, and reproducible tool for understanding how players actually play.

References

[1] "NBA Lineup Analysis on Clustered Player Tendencies: A new approach to the   positions   of basketball & modeling lineup efficiency of soft lineup aggregates", S. Kalman, J. Bosch, MIT Sloan Sports Analytics Conference, 2020

[2] Cutler, Adele, and Leo Breiman. "Archetypal Analysis." *Technometrics*, vol. 36, no. 4, 1994, pp. 338–47. *JSTOR*, https://doi.org/10.2307/1269949. Accessed 15 Nov. 2025.

[3] Canhasi, Ercan & Kononenko, Igor. (2015). Weighted Hierarchical Archetypal Analysis for Multi-document summarization. Computer Speech & Language. 37. 10.1016/j.csl.2015.11.004.

[4] G. Vinué, I. Epifanio. Archetypoid analysis for sports analytics. Data Mining and Knowledge Discovery, 31 (6), 1643-1677, 2017.