Sacramento Kings International Player Analysis

Ian Turner

October 16, 2025

Contents

Executive Summary

Objective:
The Sacramento Kings data analytics department tasked me with identifying international players who could be viable NBA targets. The goal was to use data from multiple international and NBA sources to evaluate player skill profiles, archetypes, and overall NBA readiness.

Approach:

- Combined and cleaned three JSON datasets - NBA statistics, international league data (EuroLeague, ACB, Serie A, etc.), and player metadata.
- Standardized naming conventions, merged datasets, and converted statistics into per-game and rate-based features. Combined each player's multi-league career into a single stat line using weighted averages for fair cross-season comparison.
- Developed a logistic regression model to estimate each player's probability of NBA readiness (`nba_probability`).
- Built archetype-based performance metrics (Scorer, Facilitator, Rebounder, Post Scorer, Floor Spacer, Rim Protector, Defender).
- Created an interactive Shiny web application (*Kings Player Finder*) to explore and visualize players by archetype, age, and NBA readiness.

Key Findings:

- Analysis focused on three targeted player archetypes:
  - **Two-Way Bigs:** Strong rim protection, rebounding, and post scoring.
  - **Lead Guards:** High facilitation and scoring efficiency.
  - **3&D Wings:** Excellent perimeter shooting and defensive impact.
- The top three players in each archetype were identified and visualized using 3D archetype plots.

In total, the model surfaced nine standout players across four archetypes. Among them were three **Two-Way Bigs** who combined elite rebounding and rim protection, three **Lead Guards** with high facilitation and scoring profiles, and three **3&D Wings** demonstrating strong perimeter shooting and defensive percentiles.

Recommendations:

- Prioritize scouting players with `nba_probability` > 0.5 and recent seasons (2019 or later) to ensure physical prime and current form.
- Focus on **Two-Way Bigs** and **3&D Wings**, which align with Sacramento's roster needs and modern spacing-heavy systems.
- Utilize the *Kings Player Finder* app to dynamically explore the international talent pool by archetype, age filters, and game experience.
- This analysis reinforces how valuable versatile international players can be in filling rotation gaps and complementing the Kings' existing core.

Deliverables:

- A reproducible R workflow and modeling pipeline.
- A full analytical report detailing methodology, findings, and player insights.
- An interactive Shiny web application for continued scouting use.
- All data processing and modeling were completed in R, with reproducible code included for review.

1 Introduction

Over the past several years, the NBA has become more international than ever - not only a significant surge of international players entering the league, but those players dominating at the highest level. Players like Luka Dončić, Nikola Jokić, Giannis Antetokounmpo, Shai Gilgeous-Alexander, Alperen Şengün, Victor Wembenyama, and Domantas Sabonis have proven that elite talent exists across the globe.

While scouting and drafting prospects has never been easy, the rapid growth of international basketball has made it even more complex. Evaluating college players across the U.S already requires significant travel and resources - expanding that serch to multiple continents only increases the challenge. Because of this, NBA teams have had to adapt their scouting processes to this new, globalized talent environment.

In environments where film and in-person evaluation are difficult, data becomes crucial. By leveraging detailed data from the EuroLeague, Serie A, and others, the Sacramento Kings can better identify international players who fit specific team needs. Yet even with strong box-score data, it remains difficult to truly understand a player's style. Not all centers or guards impact the game the same way - and in today's position-less, spacing focused era, understanding a player's archetype and how their skills translate to team dynamics is more important than ever.

To support the Kings' scouting and roster construction efforts, I developed a data-driven mode to evaluate international players' strengths, weaknesses, and NBA readiness. This analysis highlights which players exhibit the skill profiles most consistent with successful NBA contributors.

2 Data and Methodology

2.1 Data Sources

The data that was provided consisted of 3 JSON files, which were

- An international basketball dataset, consisting of 3,370 rows and 52 columns, with 1,473 players total in the data set. Each row consisted of a player's season with the EuroLeague, EuroCup, the Spain ACB, or the Liga A.
- A NBA basketball dataset, consisting of 1,685 rows and 55 columns, with 515 players that have played in the NBA after starting in European Leagues.

- A dataset consisting of the birth dates of every player.

The international dataset came from a different vendor than the NBA dataset, so there were multiple variables in the NBA dataset: plus-minus, calculated possessions, and plays-used. Because these variables did not appear in the international dataset, which was the primary focus of this work, these variables were not utilized.

Finally, these datasets, both the international and NBA dataset only go up to the 2020-2021 season.

2.2 Data Cleaning and Integration

First, I parsed and converted the raw JSON data (NBA, international, and player metadata) into structured CSV tables using a custom R script to extract, validate, and flatten nested JSON objects for analysis.

I then enhanced and prepared the data through the following steps:

- Standardized player names across all datasets using `toTitleCase()` for consistent formatting.
- Aggregated multi-season international data by player:
  - Summed total metrics (e.g., points, rebounds, assists).
  - Computed weighted averages for rate-based statistics using games played as weights.
- Joined international, NBA, and player metadata tables to:
  - Identify players with prior NBA experience (`nba_experience`).
  - Calculate player age as of December 31, 2021.
- Converted cumulative statistics into per-game values for comparability.
- Added each player's most recent season (`most_recent_year`) to enable recency-based filtering in later analysis.

2.3 Modeling Approach

To estimate each player's probability of NBA readiness, I trained a logistic regression model using features derived from per-game and rate-based statistics. The target variable was `nba_experience`, indicating whether a player had previously appeared in the NBA.

- Split the dataset into an 80% training and 20% testing set to evaluate model performance.
- Applied stepwise feature selection using the Akaike Information Criterion (AIC) to isolate the most predictive variables.
- Fit the logistic regression model on the training data and generated predicted probabilities (`nba_probability`) for the test set.
- Evaluated model performance using the Area Under the ROC Curve (AUC) and a confusion matrix at a 0.5 probability threshold.
- Appended the final `nba_probability` scores to the full international dataset for scouting and visualization.

Model Performance and Key Insights: The final logistic regression model achieved an AUC of **0.74** and an overall accuracy of **72.8%**, showing reliable separation between NBA and non-NBA players despite class imbalance. The most significant predictors of NBA experience were `usage_percentage_wa` ($p = 0.037$), `block_percentage_wa` ($p = 0.043$), and `total_rebounding_percentage_wa` ($p = 0.002$), all capturing traits that typically translate to NBA success—offensive involvement, rim protection, and rebounding control. Age and assist rate were marginally significant, suggesting complementary but secondary influence.

Players with higher usage rates, rebounding strength, and defensive impact were statistically more likely to have NBA experience. Importantly, several players classified as NBA-caliber but without prior NBA appearances—model "false positives"—showed strong efficiency and defensive profiles, indicating legitimate potential targets who may have been overlooked internationally.

The confusion matrix revealed a tendency toward conservative classification, with **29 true positives** and **18 false negatives**. Despite that bias, the 0.5 threshold successfully isolated high-probability players whose statistical outputs aligned with typical NBA-level production and efficiency.

Finally, I ranked international players without prior NBA experience by their predicted `nba_probability` to surface the most promising scouting candidates.

Below is a distribution of the `nba_probability` variable:

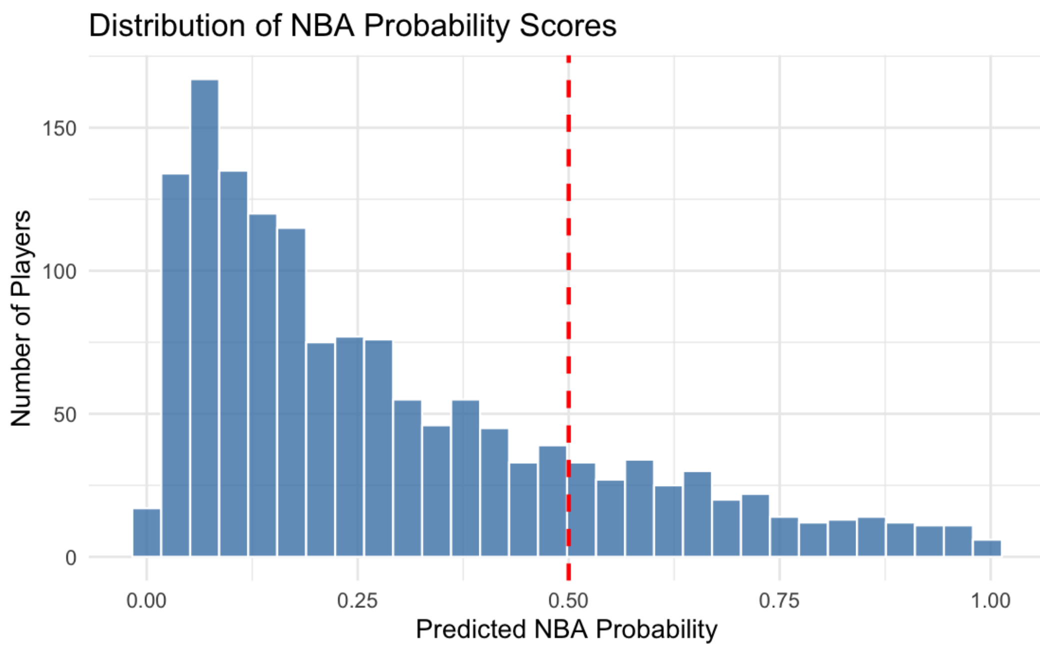


Figure 1: Distribution of predicted NBA readiness probabilities across all international players. The red dashed line marks the 0.5 threshold used for classification. Most players cluster below 0.3, while a smaller right tail represents higher-probability outliers.

2.4 Archetype Construction

While the logistic regression model quantified overall NBA readiness, it did not capture *how* players generated their value. To address this, I constructed six quantitative archetypes that summarize specific skill domains using standardized box-score data. Each archetype was derived from a separate linear model where `internal_box_plus_minus_wa` served as the response variable, allowing the coefficients from each regression to represent the relative importance of a player's underlying stats in driving overall impact. All variables were standardized as z-scores (mean = 0, standard deviation = 1) prior to modeling, ensuring comparability across metrics and positions.

- **Scorer (volume + efficiency)**

  Derived from a linear model predicting `internal_box_plus_minus_wa` using `points` and `true_shooting_percentage_wa`.

  $$\text{ScorerScore} = 1.51\, z(\text{points}) + 3.51\, z(\text{true\_shooting\_percentage\_wa})$$

- **Shot Blocker (rim deterrence without fouling)**

  Based on a model including `blocked_shots` and inverted `personal_fouls`, capturing shot-blocking activity while controlling for fouls committed.

$$\text{BlockerScore} = 1.2\, z(\text{blocked\_shots}) - 1.92\, z(\text{personal\_fouls})$$

- **Rebounder (possession control)**

  Derived from a model using `offensive_rebounds` and `defensive_rebounds`. Defensive boards were far more predictive of impact, as reflected in their larger coefficient.

$$\text{ReboundderScore} = 0.79\, z(\text{offensive\_rebounds}) + 2.16\, z(\text{defensive\_rebounds})$$

- **Facilitator (playmaking efficiency)**

  Built from a regression on `assists` and `turnover_percentage_wa`, measuring how assist creation offsets mistakes.

$$\text{FacilitatorScore} = 1.84\, z(\text{assists}) - 2.27\, z(\text{turnover\_percentage\_wa})$$

- **Floor Spacer (three-point volume + efficiency)**

  Uses `three_points_made` and `true_shooting_percentage_wa`, emphasizing both shooting output and efficiency.

$$\text{SpacerScore} = 1.25\, z(\text{three\_points\_made}) + 3.00\, z(\text{true\_shooting\_percentage\_wa})$$

- **Post Scorer (paint efficiency, volume, and foul pressure)**

  Derived from a linear model including `two_point_percentage`, `two_points_attempted`, `free_throw_percentage`, and `personal_fouls_drawn`.

$$\begin{aligned}\text{PostScore} = {} & 2.17\, z(\text{two\_point\_percentage}) + 0.24\, z(\text{two\_points\_attempted}) \\ & + 0.63\, z(\text{free\_throw\_percentage}) + 0.99\, z(\text{personal\_fouls\_drawn})\end{aligned}$$

Each composite score represents a weighted sum of standardized features, where the weights correspond to their observed contribution to overall impact as estimated by the regression models. These scores were then converted to percentiles within the dataset:

$$\text{Percentile} = 100 \times \texttt{percent\_rank}(\text{CompositeScore})$$

and mapped to descriptive performance tiers:

Percentile Range	**Tier**
95–100	Elite
85–94	Great
70–84	Good
50–69	Above Average
25–49	Below Average
0–24	Poor

Finally, for each player, each archetype was labeled with the corresponding performance tier (e.g., "Great Floor Spacer", "Good Rebounder"). These archetype percentiles form the foundation for the visual analyses in Section 3, highlighting stylistic clusters and identifying statistically distinct player profiles across international leagues.

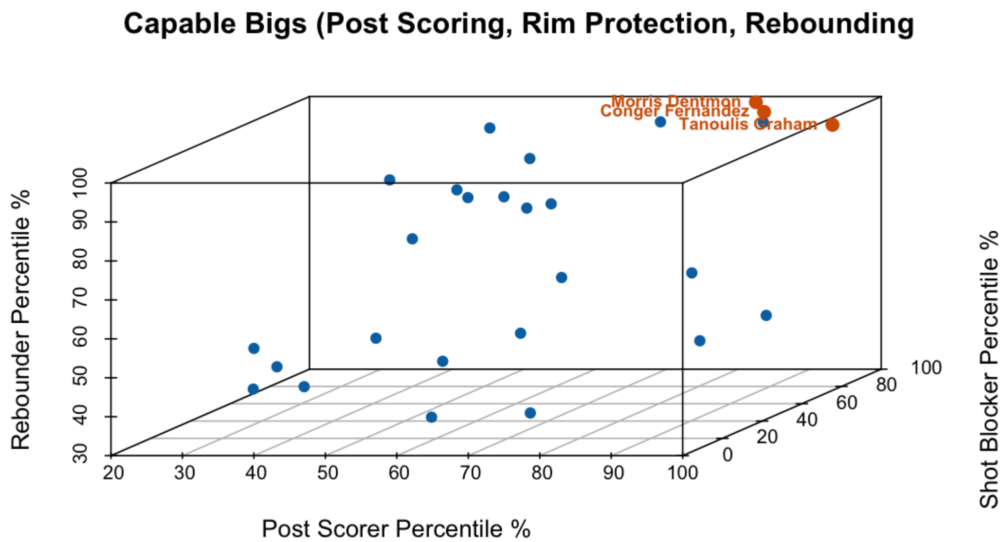
3 Results and Visualization

This section translates the statistical modeling and archetype construction into concrete scouting insights. After estimating each player's NBA readiness through the logistic regression model, the next step was to visualize how players with similar statistical profiles cluster within key archetypes. By filtering for recent, high-volume international players ($\texttt{games} \geq 10$, $\texttt{minutes} \geq 15$, age ≤ 33, and season ≥ 2019) and excluding anyone with prior NBA experience, the analysis highlights potential acquisition targets whose statistical impact most closely resembles that of NBA-caliber players ($\texttt{nba\_probability} > 0.5$).

The following subsections present 3-dimensional archetype visualizations. Each plot highlights the top three candidates in orange, with axes representing the percentile values of the relevant skill dimensions. Tables accompany each plot, ranking all qualified players by their composite archetype scores.

3.1 Capable Bigs (Post Scoring, Rim Protection, Rebounding)

This visualization highlights versatile frontcourt players capable of contributing through paint scoring, rim protection, and rebounding. By averaging each player's percentiles across the post_scorer, shot_blocker, and rebounder archetypes, a composite “balance” score was derived. The 3D scatter plot maps these dimensions simultaneously, with the top three candidates—marked in orange—representing the most balanced, high-impact bigs lacking prior NBA experience.



Below is the basic statistics of the 3 most well rounded big men, along with their statistics and archetype ratings.

Table 1: Top Centers By Composite Frontcourt Score (n = 3)

Name	Age	PTS	REB	AST	G	Scorer	Rebounder	Shot Blocker
Tanoulis Graham	32	12.9	5.8	1.1	186	Great	Great	Elite
Morris Dentmon	25	11.3	8.7	0.7	12	Above Avg	Elite	Elite
Conger Fernandez	25	15.1	6.8	0.8	10	Good	Elite	Elite

3.2 Lead Guards (Scoring, Facilitation, NBA Probability)

This view isolates perimeter creators who pair shot–making with decision–making. A composite *guard score* averages the scorer_percentile and facilitator_percentile. The 3D plot maps Scorer and Facilitator percentiles against NBA probability, with the top three candidates highlighted in orange.

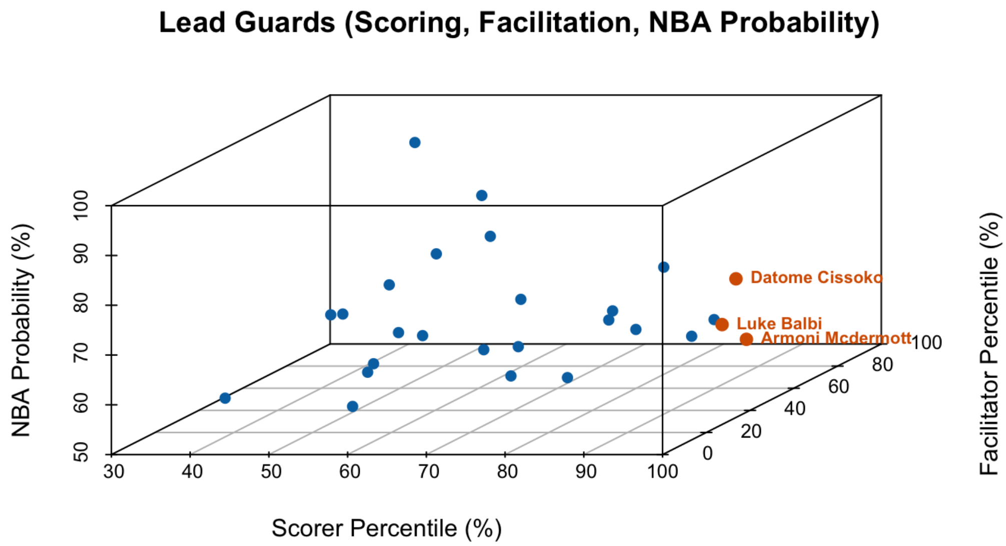


Figure 2: 3D archetype visualization of lead guards showing Scorer and Facilitator percentiles versus NBA probability. Orange points denote the top three guards by composite score (`guard_score`).

Table 2: Top Lead Guards by Composite Guard Score (n = 3)

Name	**Age**	**PTS**	**REB**	**AST**	**G**	**Scorer**	**Facilitator**	**Floor Spacer**
Armoni Mcdermott	27	15.5	2.8	3.7	19	Good	Elite	Above Avg
Datome Cissoko	32	13.1	3.0	4.3	163	Good	Great	Great
Luke Balbi	28	14.1	2.0	3.5	147	Good	Great	Great

3.3 3&D Wings (Shooting, Defense, NBA Probability)

The 3&D archetype captures players who provide perimeter shooting and defensive versatility—traits crucial for scalable NBA role players. Among international players without NBA experience, the analysis reveals that few exhibit elite balance in both shooting and defensive metrics.

While several players demonstrate strong individual skills, the top composite scores peak near the 80^{th} percentile, suggesting that true two-way perimeter contributors are scarce in the international talent pool.

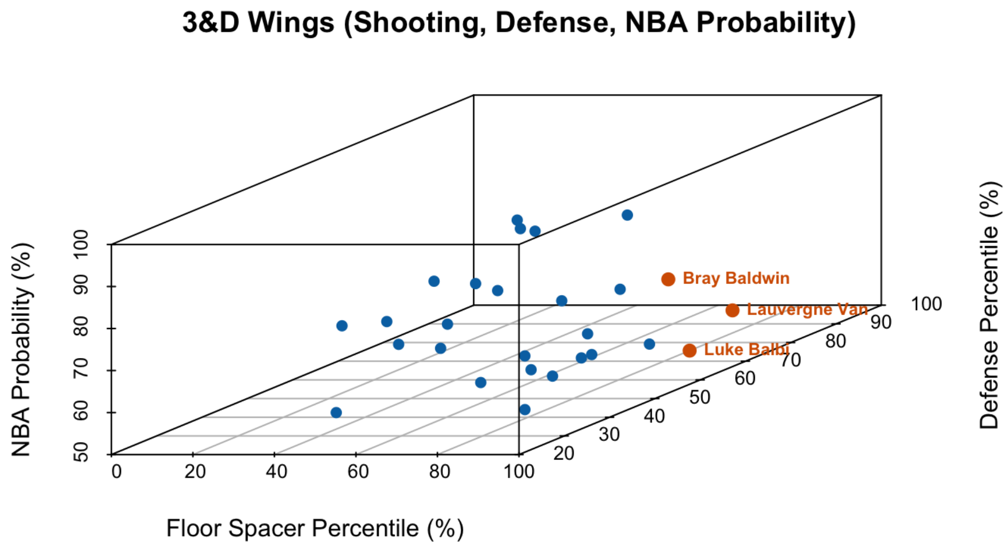


Figure 3: 3D archetype visualization of international perimeter players showing percentiles for three-point shooting efficiency, defensive impact (steal% + block%), and overall NBA probability. Orange points indicate the top three candidates by composite 3&D score, combining `floor_spacer_percentile` and `defense_percentile`.

Table 3: Top 3 3&D Wing Candidates by Shooting and Defense Composite Score

Name	Age	PTS	AST	G	Floor Spacer	Defense	3PT%
Lauvergne Van	28	13.7	2.5	61	Great	Good	36.8%
Luke Balbi	28	14.1	3.5	147	Great	Above Avg	37.9%
Bray Baldwin	31	7.4	1.5	262	Above Avg	Good	37.3%

3.4 Archetype Translation to NBA Success

To contextualize the logistic model's output, I examined how different archetypes correspond to actual NBA experience among international players. Across the full dataset, approximately **29%** of players have appeared in the NBA. By isolating only the top 30% performers within each archetype (those rated "Good" or higher), the analysis below compares which skill profiles are most frequently represented among players who reached the league. This provides a direct measure of which archetypes are historically most valued by NBA teams.

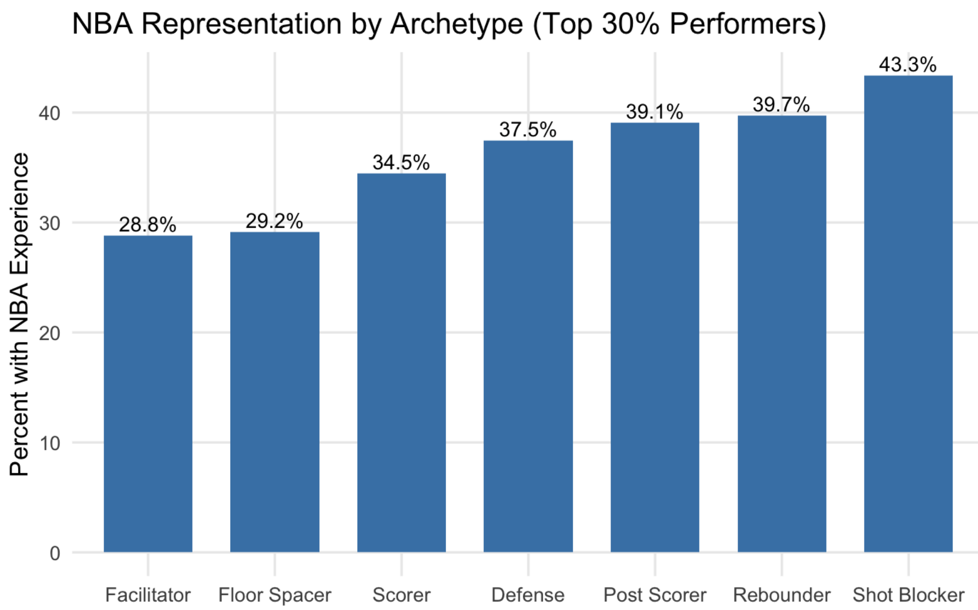


Figure 4: NBA representation by archetype, showing the percentage of "Good" or better players in each category with prior NBA experience. Archetypes emphasizing rebounding, rim protection, and interior impact have the highest historical translation to NBA play, while facilitation and shooting archetypes appear less represented among international players with NBA experience.

4 Discussion and Future Applications

4.1 Key Takeaways

This analysis demonstrates how data-driven player profiling can support international scouting by identifying players whose statistical performance mirrors that of NBA contributors. The following key insights emerged:

- The logistic regression model achieved an AUC of **0.74** and accuracy of **72.8%**, effectively distinguishing NBA-caliber players from those without league experience.
- False positives—players classified as NBA-ready but without prior NBA experience—represent high-value scouting opportunities, indicating statistical profiles that align closely with NBA-level talent.
- Archetype-based analysis provided role-specific context, revealing that interior archetypes such as **Rebounders**, **Post Scorers**, and **Shot Blockers** were most predictive of NBA translation, while perimeter-oriented archetypes exhibited greater variance.

- Among all players analyzed, approximately **29%** had prior NBA experience, reinforcing the depth of international talent and the opportunity for data-informed discovery.

Across the three role categories analyzed—**Capable Bigs**, **Lead Guards**, and **3&D Wings**—the top-ranked players exemplified balanced statistical profiles within their archetypes, validating the framework's ability to distinguish translatable player types across positions. These examples demonstrate how quantitative archetype modeling can augment traditional scouting by efficiently surfacing international players whose performance metrics align with established NBA success patterns.

4.2 Limitations

While the modeling framework produced meaningful insights, several limitations should be acknowledged when interpreting the results:

- **Data Coverage:** The dataset extends only through the 2020–2021 season, which may not reflect current player form or development trajectories. Players who improved significantly after that period would be undervalued by the model.
- **League Context:** The international datasets vary in competitiveness, pace, and style of play. Without league-strength normalization, players in weaker leagues may appear statistically inflated relative to those in top-tier competitions such as the EuroLeague.
- **Feature Scope:** The model relies exclusively on box-score and rate-based statistics. It does not incorporate tracking data (e.g., shot quality, spacing impact, defensive positioning) that would provide a more complete assessment of translatability.
- **Missing Player Attributes:** The absence of variables such as height, weight, and position limited the model's ability to contextualize statistical output. For example, centers could not be distinguished from forwards when evaluating rebounding or rim protection efficiency, reducing positional interpretability.
- **Limited Defensive Detail:** Several potentially valuable indicators—`deflections`, `screen assists`, `second chance points`, `loose balls recovered`, and `charges drawn` —contained almost exclusively zeros and could not be analyzed. As a result, defensive evaluation relied primarily on steals and blocks, a basic but incomplete proxy for defensive impact.
- **Model Simplicity:** Both the logistic regression and the archetype regressions assume linear relationships between inputs and outcomes. Real-world player value is often nonlinear and context-dependent, particularly for multi-skill players.

- **Role Ambiguity:** Players frequently contribute across multiple archetypes. Although percentiles mitigate overlap, the framework still assigns roles independently rather than modeling positional fluidity or stylistic interaction.

Despite these constraints, the system provides a reproducible and interpretable foundation for quantifying international player impact. Future iterations incorporating more recent seasons, positional data, contextual weighting, and event-level metrics could further refine accuracy and scouting relevance.

4.3 Web Application Implementation

To make the results more accessible for scouting and coaching staff, I developed an interactive **Shiny web application**, available at:

https://ianturner25.shinyapps.io/kings-player-finder/

The tool allows users to explore the international player database by archetype, age, and predicted NBA probability in real time. It supports filtering by archetype tiers (e.g., Elite Rebounders, Great Floor Spacers), visualizes player percentiles through interactive 3D plots, and provides on-demand statistical summaries.

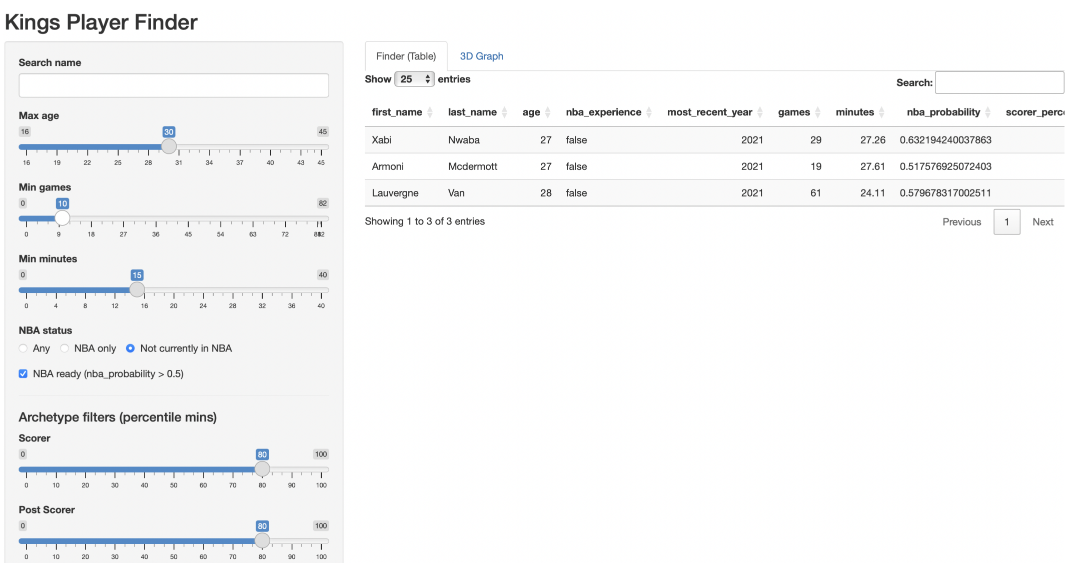


Figure 5: Kings Player Finder web application — user interface with filtering controls.

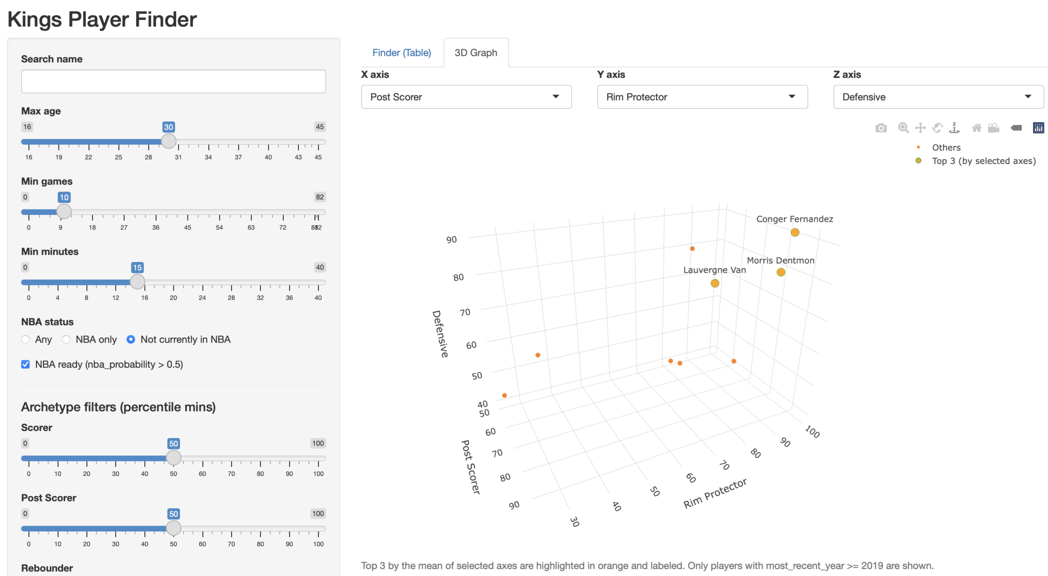


Figure 6: Example visualization output highlighting player archetype balance and NBA probability.

The web application transforms static model outputs into a dynamic scouting tool, enabling real-time exploration of the international player landscape. It can be extended to include updated seasons, positional filters, or integration with internal scouting databases.

4.4 Future Applications

The framework and analytical pipeline developed in this project can be expanded to support ongoing scouting, player development, and strategic decision-making. Several future directions are outlined below:

- **Data Expansion:** Incorporate additional international seasons (2021–present) and integrate player metadata such as height, position, and wingspan to better contextualize performance. These attributes would improve positional comparisons—especially for centers and forwards—and strengthen archetype classification.

- **Enhanced Defensive Metrics:** Include non-traditional box-score variables such as deflections, screen assists, second chance points, loose balls recovered, and charges drawn as data quality improves. These inputs would provide a more nuanced view of defensive impact beyond steals and blocks.

- **Automated Model Updating:** Develop a scheduled pipeline to automatically pull and clean new data, retrain the model, and refresh the *Kings Player Finder* web application to maintain relevance through each international season.

- **League Normalization:** Create standardized strength coefficients for each league (e.g., EuroLeague, ACB, Serie A, NBL) to normalize player production across competitive environments.

- **Integration with Internal Systems:** Link the model and web application with the Kings' internal scouting databases or film systems to combine statistical and qualitative evaluations into a unified interface.

- **Extended Use Cases:** Apply the archetype and probability framework to domestic G League or Summer League data to benchmark international prospects against fringe NBA players and identify stylistic "lookalikes."

By continuing to expand this system, the Kings can build a scalable, data-driven scouting infrastructure that complements traditional evaluation methods. This integration of analytics, visualization, and interactive tooling represents a step toward an adaptive model of international talent identification.

4.5 Conclusion

This project demonstrates the value of integrating statistical modeling, archetype construction, and interactive visualization to modernize international scouting. By combining logistic regression–based NBA readiness estimation with archetype-specific performance analysis, the framework provided both quantitative depth and practical scouting utility.

The results highlighted that approximately 29% of international players in the dataset have NBA experience, with frontcourt archetypes—particularly rebounders, rim protectors, and post scorers—showing the strongest historical translation to league success. At the same time, several non–NBA players surfaced as statistically comparable to established professionals, underscoring the system's ability to uncover under-scouted prospects.

Through the *Kings Player Finder* web application, these insights were transformed into a dynamic tool capable of supporting data-informed decision-making in real time. Future iterations that incorporate positional, contextual, and tracking data will further refine the predictive accuracy and scouting relevance of this approach.

Overall, this work provides a scalable foundation for the Sacramento Kings to identify, evaluate, and prioritize international talent more efficiently—bridging quantitative analytics with traditional basketball intuition to strengthen long-term roster construction.