



UNIVERSIDAD DE ALMERÍA

GRADO EN MATEMÁTICAS – ANÁLISIS DE DATOS

Informe del Valor y Potencial de Jugadores NBA

Desde el Draft hasta el All Star

Autor: Ángel Ramos Ortiz

2 de junio de 2025

Índice

1. Resumen	2
2. Estudio Salarial	3
3. Estudio de Perfiles de Juego NBA	10
4. Estudio del All-Star	15
5. Estudio del Draft	18
6. Conclusión	22

1 Resumen

La NBA (*National Basketball Association*) es la liga profesional de baloncesto más importante de Estados Unidos y las más reconocida a nivel mundial. Está compuesta por 30 equipos, conocidos como franquicias, separados geográficamente en conferencias Este y Oeste. A lo largo de su historia, la NBA se ha caracterizado por una fuerte tradición en la recolección de datos y el uso de analítica deportiva, lo que la convierte en un entorno ideal para la exploración estadística. Además, elementos únicos como el sistema de draft, el All-Star Game y la propia complejidad táctica del baloncesto, permiten desarrollar una amplia variedad de análisis cuantitativos con un alto valor interpretativo.

Este informe tiene como objetivo abordar las siguientes cuestiones mediante el uso de estadísticas deportivas:

- Determinar qué factores de juego influyen más en el salario de los jugadores de la NBA.
- Detectar posibles jugadores sobrevalorados o infravalorados salarialmente en cuanto a rendimiento deportivo.
- Identificar arquetipos de juego a partir de estadísticas individuales.
- Predecir qué jugadores formarán parte del *All-Star* en una temporada determinada.
- Analizar el sistema universitario y el draft de la NBA, identificando universidades clave y prediciendo el éxito profesional.

2 Estudio Salarial

En esta sección se presenta un análisis centrado en el estudio y la predicción del salario de los jugadores de la NBA, utilizando exclusivamente estadísticas de juego y métricas deportivas. Este enfoque puede resultar de especial interés para los representantes de franquicias NBA que deseen establecer una política salarial basada en el rendimiento deportivo objetivo de cada jugador durante la temporada.

La predicción del salario es un reto complejo, ya que este suele estar influido por múltiples factores ajenos al desempeño en cancha, como el marketing, la popularidad, o decisiones estratégicas del equipo. Por ello, se ha planteado un análisis riguroso, enfocado únicamente en variables deportivas, para valorar en qué medida estas pueden explicar o predecir la remuneración de los jugadores.

El primer paso del análisis consistió en ajustar todos los salarios a una misma escala económica, eliminando el efecto de la inflación a lo largo del tiempo. Esto permite comparar el valor económico de los jugadores en base a la economía deportiva actual.

A continuación, se representa la distribución de los salarios brutos ajustados por inflación:

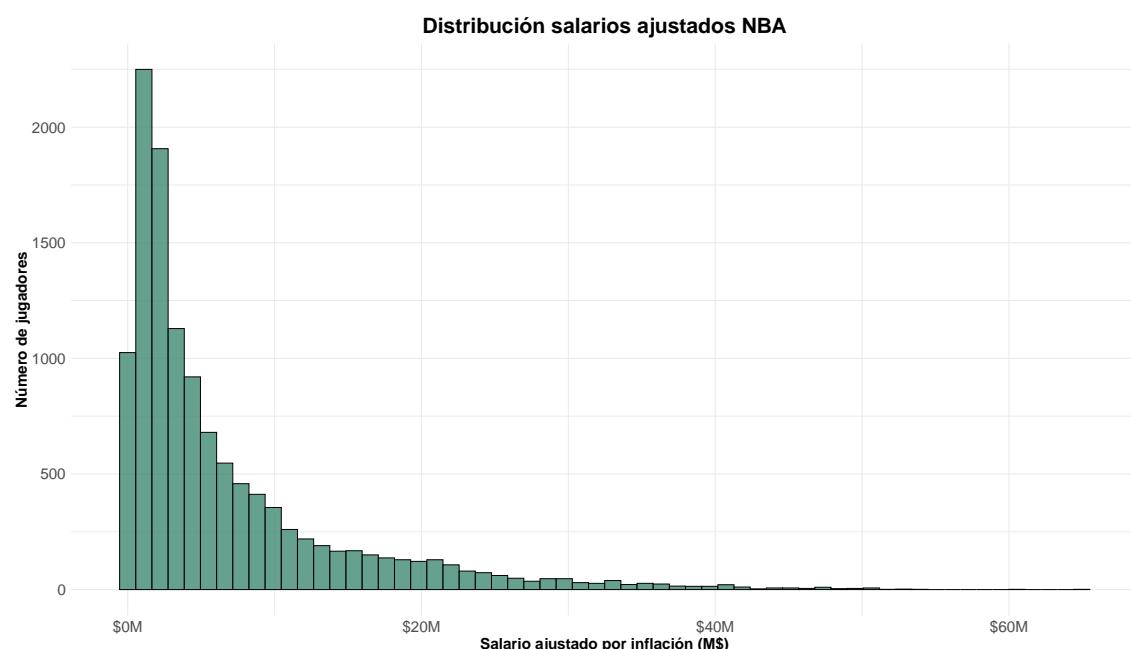


Figura 1: Histograma de salarios brutos ajustados por inflación en la NBA (1997-2024)

Como se observa en el histograma, la mayoría de los jugadores se concentran en rangos salariales bajos, especialmente por debajo de los 20 millones de dólares. A medida que el salario aumenta, el número de jugadores en esos tramos disminuye significativamente, aunque se mantienen algunos casos extremos que alcanzan hasta los 60 millones de dólares. Aunque son pocos, existen jugadores con remuneraciones muy por encima de la media.

En términos generales, el salario promedio en el periodo analizado ha sido de aproximadamente 6.7 millones de dólares. La mediana salarial se sitúa en 3.6 millones de dólares, lo que divide a los jugadores en dos grupos de igual tamaño: la mitad gana por encima de esa cifra y la otra mitad por debajo.

Informe de Resultados

Como cabía esperar, la relación entre las estadísticas individuales y el salario no sigue un patrón lineal sencillo. En el siguiente gráfico se muestran comparaciones entre el salario y cuatro métricas clave de rendimiento. Se puede observar cómo, aunque se detectan ciertas tendencias, las líneas rectas ajustadas a los datos explican solo una parte limitada de la variabilidad salarial.

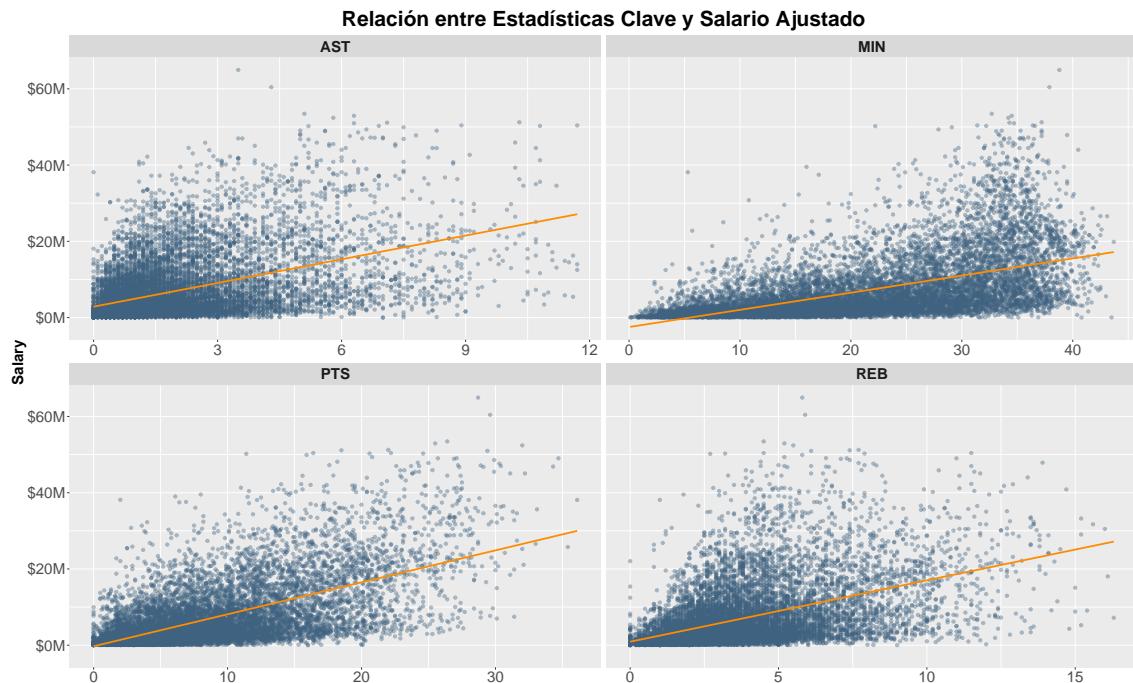


Figura 2

La experiencia en la liga juega un papel fundamental en el salario de un jugador. La siguiente figura es buena prueba de ello:

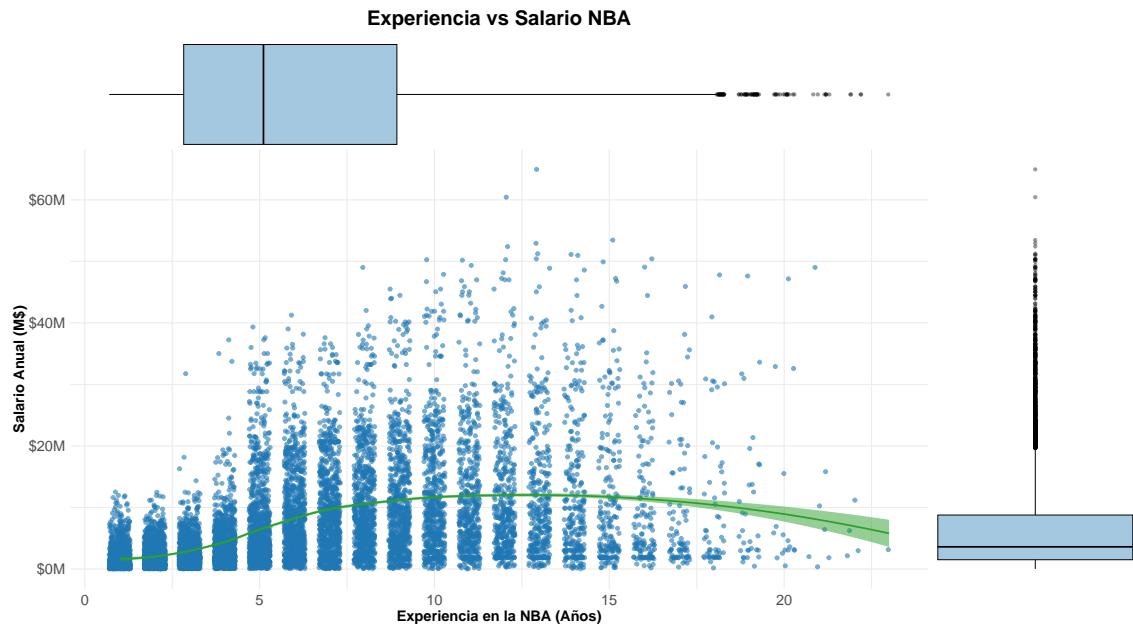


Figura 3

Informe de Resultados

Se observa que los jugadores novatos en la liga tienden a percibir salarios significativamente más bajos en comparación con aquellos que ya cuentan con una carrera consolidada, especialmente a partir de los cinco años de experiencia. Sin embargo, en casos más excepcionales, cuando un jugador ha superado los 20 años en la liga, es habitual observar una disminución en su salario, posiblemente relacionada con un menor protagonismo en pista o una fase final de su carrera.

En la figura se representan diagramas de caja marginales que permiten visualizar la experiencia en la NBA y el salario por separado. Entre otros aspectos, se destaca que la mediana de experiencia de los jugadores en el conjunto de datos es de aproximadamente cinco años, lo que indica una presencia notable de jugadores jóvenes en cada temporada.

Surgen a partir de aquí algunas preguntas de interés. Una de ellas es si existe una diferencia salarial significativa entre las distintas posiciones en el baloncesto profesional estadounidense. Para este análisis se han agrupado las posiciones tradicionales en tres grandes categorías:

- **Exteriores (bases y escoltas)**
- **Alas (Aleros y Alas-Pívots)**
- **Pívots**

Todas estas posiciones desempeñan roles fundamentales dentro de un equipo, por lo que cabría esperar una distribución salarial relativamente equilibrada entre ellas.

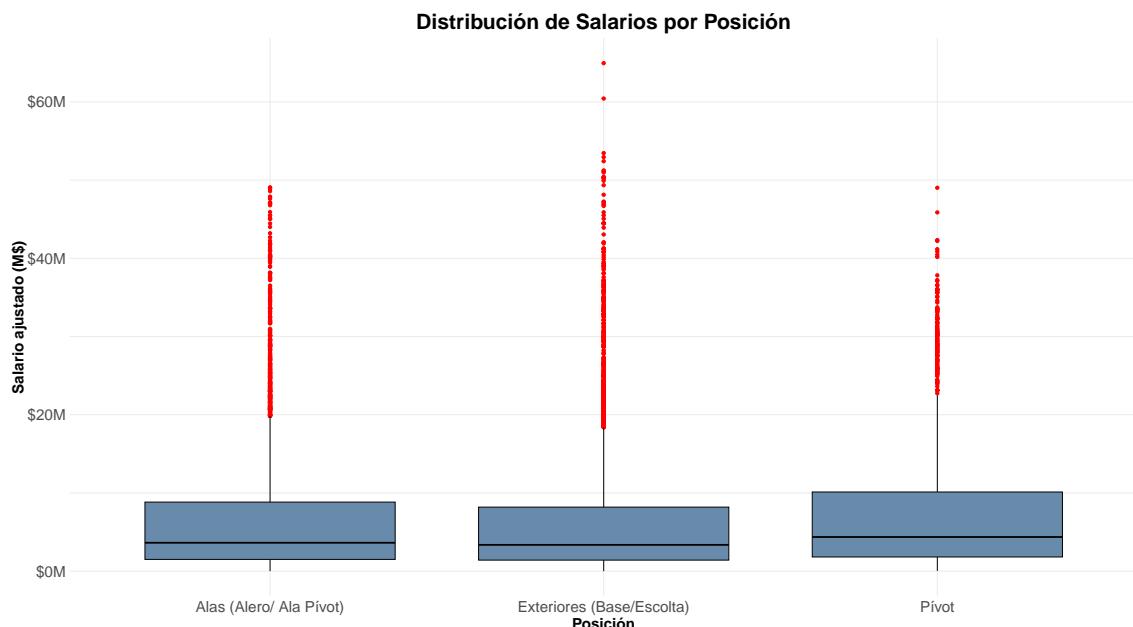


Figura 4

En efecto, la distribución del salario de las diferentes posiciones varía muy poco. Parece que el pívot promedio percibe un poco más que el resto de posiciones. Sin embargo, en aquellos *outliers* salariales formados por superestrellas de la liga, los jugadores exteriores son los que han conseguido los salarios más altos a lo largo de estos años, eso refleja su mayor visibilidad e impacto en ciertos casos.

Informe de Resultados

Dado que la posición no muestra ser un factor claramente diferenciador en términos salariales y añade complejidad al análisis sin aportar una mejora sustancial a la capacidad predictiva, se ha decidido no incluirla como variable en los modelos desarrollados.

Finalmente, se plantea una cuestión especialmente relevante desde una perspectiva de eficiencia económica: ¿existen jugadores mal valorados salarialmente en relación con su rendimiento deportivo?

Teniendo en cuenta los factores contractuales, es razonable pensar que sí. Para explorar este aspecto, se ha comparado el rendimiento deportivo —medido mediante la métrica NBA Fantasy Points— con el salario de los 25 jugadores mejor pagados en la temporada 2018-19. Esta métrica tiene en cuenta múltiples estadísticas acumuladas a lo largo de la temporada, ofreciendo una aproximación bastante representativa del aporte total de un jugador a su equipo.

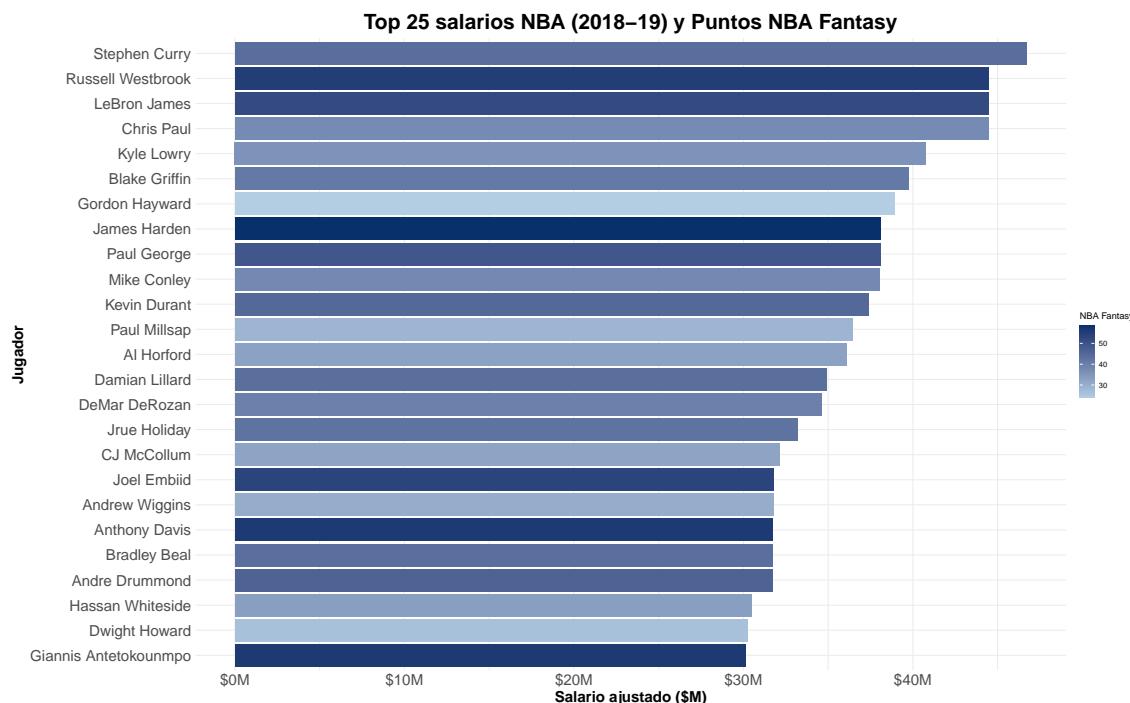


Figura 5

Se detectan casos de jugadores con salarios notablemente altos en relación con su producción deportiva. Por ejemplo, Gordon Hayward aparece como uno de los jugadores con salario más inflado, mientras que Giannis Antetokounmpo, pese a ofrecer un rendimiento excepcional, ocupaba únicamente el puesto 25 en la escala salarial, cuando sus méritos lo situarían con claridad entre los cinco mejores remunerados de esa temporada.

A continuación, se analiza la evolución temporal de la media salarial en la NBA. Aunque todos los valores han sido ajustados por inflación a niveles económicos actuales, esto no implica que la media se haya mantenido constante a lo largo de los años.

Informe de Resultados

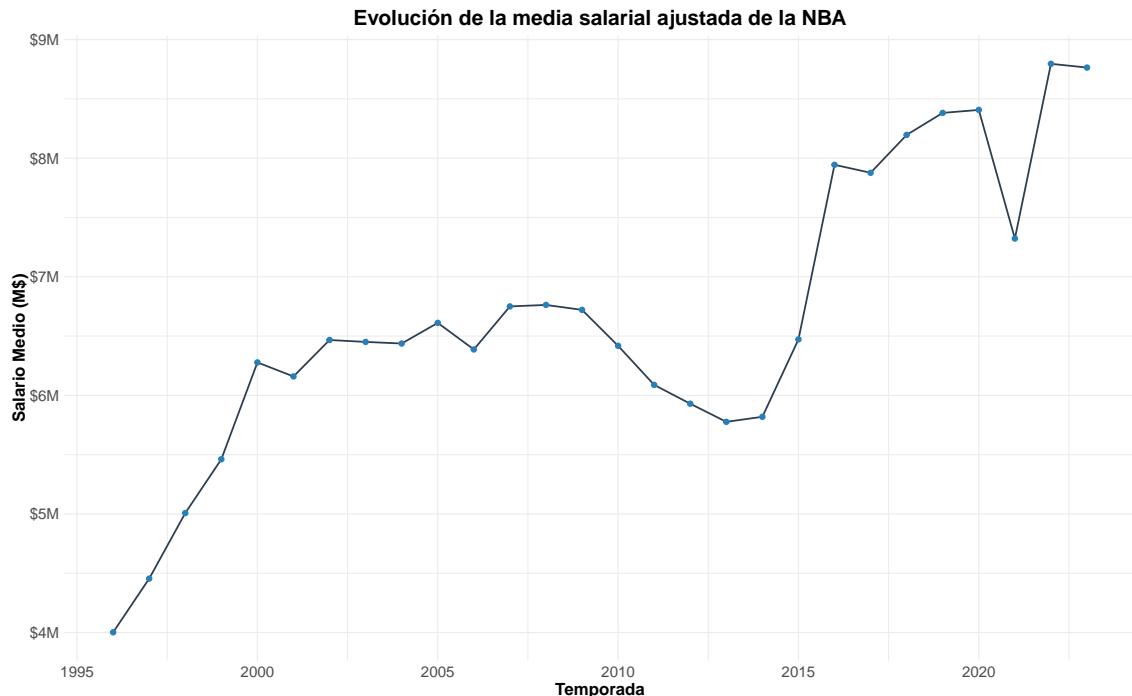


Figura 6

Tal y como muestra la figura, la media salarial presenta una clara tendencia ascendente, aunque con algunas caídas puntuales. Estas variaciones no se explican únicamente por la inflación, sino también por otros factores macroeconómicos y estructurales propios de la liga. Algunos de los momentos clave que explican los cambios más notorios son:

- En 2016 la NBA renovó su contrato televisivo con Disney (ESPN) Y TNT con el acuerdo más grande de la historia de la liga, de ahí el crecimiento tan grande que se dio en 2016. Como dato, el acuerdo fue firmado durante 9 temporadas y la NBA lo renovará este mismo año con otro contrato récord que sin duda volverá a afectar a los salarios de los jugadores.
- En 2020 y 2021 la pandemia hizo que los salarios bajaran repentinamente aunque luego más tarde remontarían a máximos históricos.

Este aspecto es clave para el modelado, pues es necesario que el modelo que se use capte las diferencias salariales entre temporadas. Sobre todo debe captar cambios bruscos como la pandemia o el acuerdo televisivo si se quiere que modele bien.

Para predecir el salario usaremos XGBoost (Extreme Gradient Boosting), un algoritmo basado en árboles de decisión que utiliza la técnica de *boosting*.

Entrenado y refinado usando cerca de unos 10000 jugadores, el modelo ha obtenido unas métricas de

Las métricas obtenidas aplicadas a nuestro conjunto *test* son:

- R^2 : 0.754. Lo que indica que el modelo explica el 75 % de la varianza en los datos.
- Error medio absoluto ajustado: \$2870036

Informe de Resultados

El error medio cometido puede parecer grande, pero teniendo en cuenta que es error medio y que el modelo se equivocará sobre todo en contratos muy inflados o infravalorados, así como la escala que tienen los salarios de la NBA, el modelo rendirá realmente bien y ayudará a detectar jugadores mal remunerados.

En cuanto a la importancia de las variables predictoras se obtiene lo esperado. La experiencia es la variable más importante para el modelo, seguida de los puntos por partido. La temporada adquiere un peso notable en las decisiones del modelo, lo que indica que está captando muy bien cómo va cambiando el salario con el paso de las temporadas. Luego siguen estadísticas como minutos, rebotes, asistencias y acierto. A continuación se muestran las diferentes variables y su aporte:

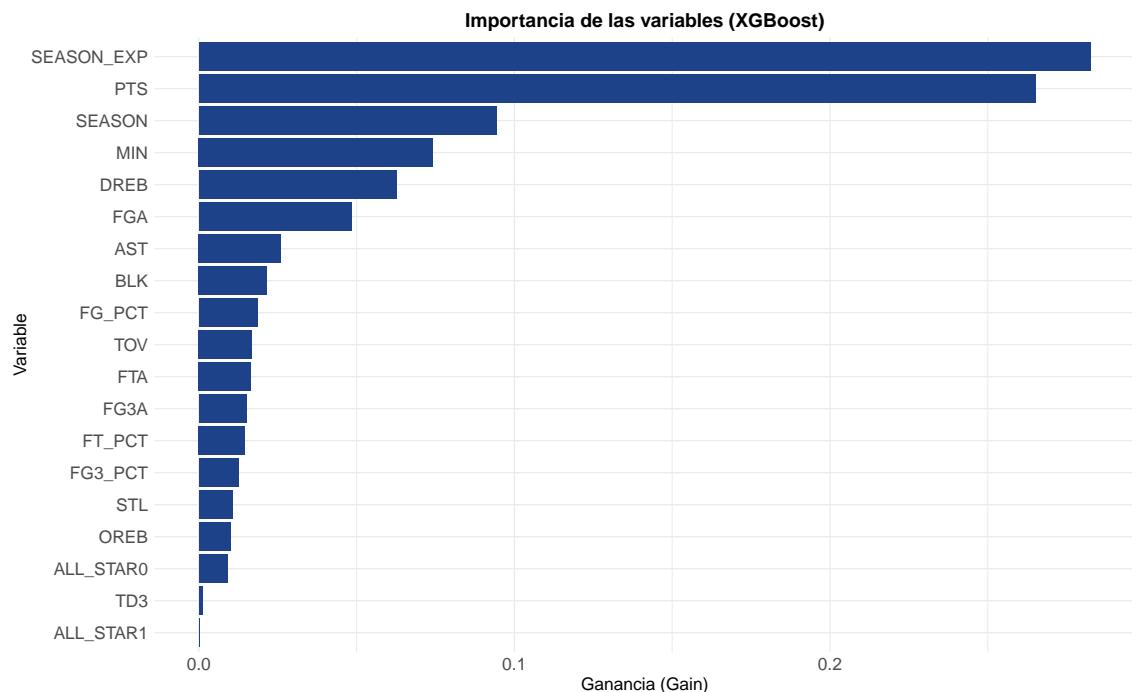


Figura 7

En lo que sigue se muestran algunos ejemplos concretos de **jugadores del conjunto test** (formado por cerca de unos mil jugadores con los que el modelo no ha sido entrenado) para ver cómo predice su salario el modelo:

- Para Stephen Curry el modelo predice con un error de solamente \$81066. Se trata de la temporada 2012-13 en la que ganaba 5.42M de dólares, es decir, el error es del 0.015 %. Se trata de uno de los casos en los que el modelo predice mejor.
- En la 2022-23 un joven Zion Williamson ya 3 años en la liga, firma su nuevo contrato y pasa a cobrar unos 14.4M de dólares. El modelo falla unos 283 mil dólares, una aproximación realmente buena teniendo en cuenta el cambio contractual del jugador. El error es del 0.020 %.
- El modelo falla estrepitosamente en algunos casos como el de Mike Conley. Conley cobraba en la temporada 2017-18 un salario de 36.5 millones de dólares. El modelo predice unos 14 millones de dólares menos. La razón es clara, pese a ser un más que buen jugador, el contrato estaba claramente inflado. A día de hoy cobra 9 millones de dólares. Es un claro caso de sobrevaloración que el modelo ayuda a detectar.

- Jamal Crawford en la temporada 2003-04 percibía un salario de 4.4M de dólares. El modelo predice que cobraría unos 11 millones. Al consultar las estadísticas se puede ver que su experiencia en la liga era de unos 4 años y que su media eran 17 puntos por partido. Teniendo en cuenta la época, seguramente se trataba de uno de los mejores bases de la liga. Este es un caso de jugador infravalorado. Efectivamente, se puede consultar que al año siguiente su contrato subió a los 9 millones de dólares y a las dos temporadas a los 11 millones estimados por el modelo.
- Para Shaquille O'Neal en la 2008-09 el modelo erra unos 4.3M de dólares. Esta cantidad puede parecer muy grande, pero O'Neal cobraba unos 30.02M de dólares y era un *All-Star*. Nótese que el error no llega al 15 %.

El modelo desarrollado se basa exclusivamente en el rendimiento deportivo de los jugadores, y aun así ha logrado captar de forma bastante precisa el valor relativo de la mayoría de ellos. Este logro es especialmente destacable si consideramos que existen múltiples factores ajenos al desempeño en cancha que influyen significativamente en los salarios y que no han sido incluidos en el análisis, entre ellos:

- La capacidad de generar ingresos por marketing o atraer afición.
- La evolución del tope salarial impuesto por la liga cada temporada.
- La existencia de contratos de patrocinio y acuerdos comerciales.
- La influencia de agentes y negociaciones contractuales.

La situación estratégica de la franquicia en la que milita el jugador —por ejemplo, equipos en proceso de reconstrucción tienden a experimentar una mayor rotación de plantilla, lo que afecta a las expectativas salariales¹.

A pesar de no incorporar estos elementos externos, el modelo ha mostrado un comportamiento robusto y consistente, lo que demuestra que el rendimiento estadístico puro puede ser una base fiable para estimar el valor económico de un jugador en términos comparativos.

En otras palabras, este modelo proporciona una herramienta objetiva y transparente para evaluar salarios justos a partir de datos medibles. Puede ser de gran utilidad como referencia interna para negociaciones, *scouting* o diseño de políticas salariales más alineadas con el rendimiento real en pista.

¹En la NBA muchas franquicias atraviesan ciclos de alto rendimiento seguidos de etapas de reconstrucción, donde la rotación de jugadores es alta y se prioriza la flexibilidad financiera.

3 Estudio de Perfiles de Juego NBA

Dentro de la NBA hay diferentes estilos de juego: jugadores hablidosos y regateadores, tiradores de tres, interiores defensivos, aleros polivalentes, etc.

Distinguiendo entre posiciones ya se pueden encontrar diferencias evidentes:

- **Exteriores (bases y escoltas):** Son los encargados de dirigir el juego y organizar el ataque. Suelen ser rápidos, buenos manejando el balón y tienen buena visión de juego y tiro exterior. El base es el cerebro del equipo y el escolta suele ser un anotador ágil.
- **Alas (aleros y ala-pívots):** Juegan en posiciones intermedias entre exteriores y pívots. Son versátiles: deben saber tirar, defender, moverse sin balón y ayudar tanto en el rebote como en el contraataque. Mezclan fuerza y agilidad.
- **Pívots (pívots):** Se ubican cerca del aro. Su función principal es dominar en la pintura, anotar cerca del aro, capturar rebotes y proteger la canasta. Suelen ser los jugadores más altos y fuertes del equipo.

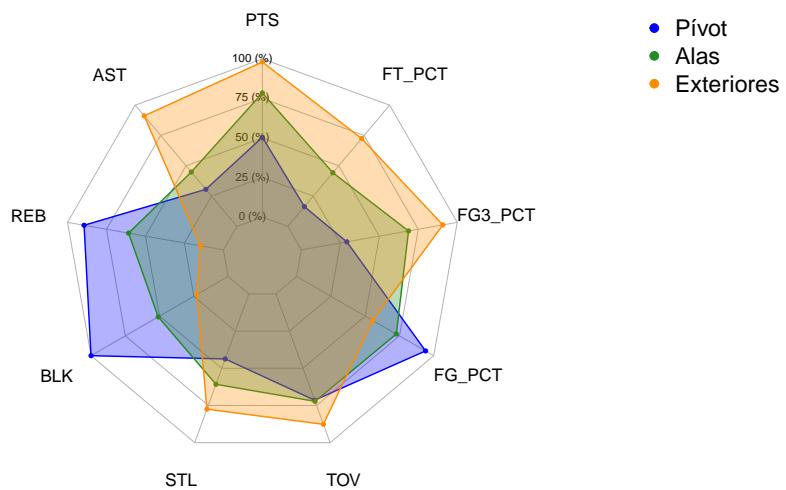


Figura 8: Gráfica de radar o araña que muestra las diferencias de juego entre las diferentes posiciones.

La gráfica permite ver rápidamente que:

- **Exteriores:** dominan el acierto en tiros libres y de tres, las asistencias y la anotación, así como los robos. Son los jugadores más débiles en rebote y tapones, además de ser los que más pierden el balón.
- **Alas:** sin llegar a ser los mejores en ningún aspecto (en promedio) ni los peores en nada (en promedio), los alas son sin duda los jugadores más completos. Constituyen un híbrido ideal para transiciones y jugadas donde se requieren diferentes habilidades.

Informe de Resultados

- **Pívots:** los pívots dominan los registros de rebotes y tapones como era de esperar. A su vez tienen un mayor porcentaje de tiros de campo debido a que sus tiros suelen ser cerca del aro. No obstante, fallan más que el resto en triples y tiros libres. Además, son los que menos asistencias y robos producen.

NOTA: la gráfica de radar es una figura muy criticada por diversos motivos en la comunidad estadística ya que puede llevar a confusiones por sus cambios de escalas y la visualización radial aunque destaca su fuerte uso en la analítica deportiva.

A continuación se visualizan los datos anteriores pero con una visualización más adecuada para distinguir las magnitudes de las diferentes variables:

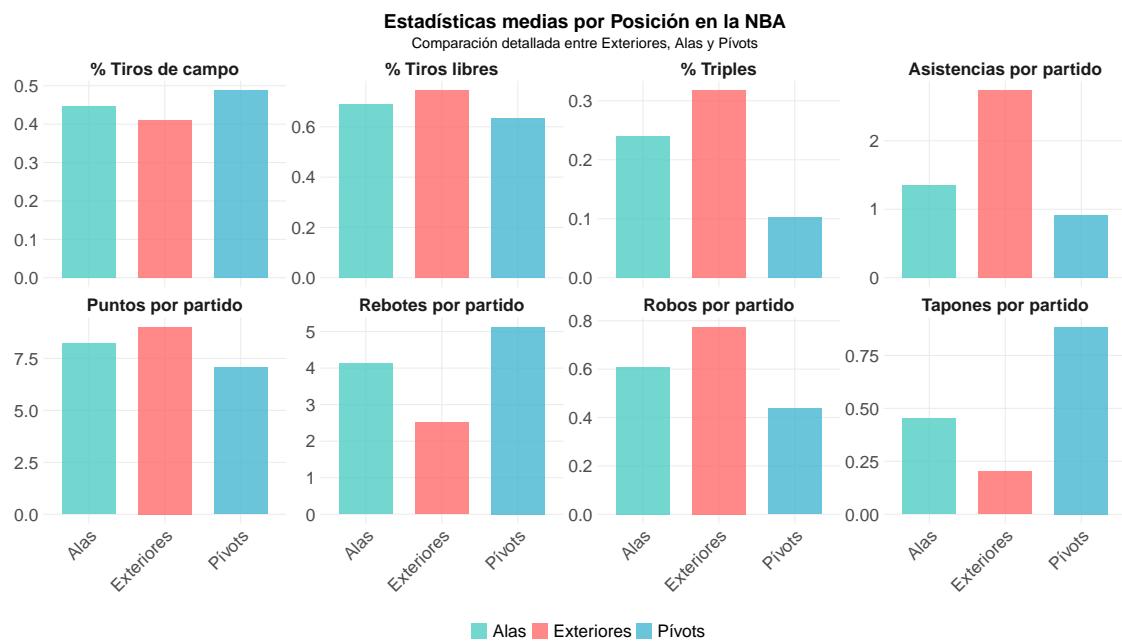


Figura 9

Los jugadores de la NBA poseen unas características muy diferentes entre sí. Simplemente por posiciones ha sido posible evidenciar ciertas tendencias a estilos de juego diferentes. No obstante, esto ha sido observado para la media de las posiciones durante unas 27 temporadas.

En la NBA moderna los jugadores asumen roles diferentes e incluso a veces inesperados para sus posiciones. Esto se debe a la transformación táctica que sufren los deportes con el paso del tiempo y que obliga a la aparición de perfiles de jugadores innovadores y diferentes. Un ejemplo de ello es el pívot serbio de los Denver Nuggets, Nikola Jokić. Pese a ser el conocido big man del equipo, fuerte en la pintura y en defensa, Jokić ha demostrado ser uno de los pívots con mejor pase de la historia y desde la línea de tres ocasionalmente prueba suerte obteniendo muy buenos resultados, algo impensable para pívots de principios de siglo como Shaquille O’Neal. Esta polivalencia, unida a su dominio, le valió a los Nuggets el primer anillo de su historia hace un par de años, pues Jokić además de ser impredecible fue elegido MVP de las finales.

Por ejemplos como este último tiene todo el sentido del mundo recurrir al modelado estadístico para identificar perfiles de juego. En este caso se recurirá a técnicas de aprendizaje no supervisado y reducción de la dimensionalidad sobre las estadísticas de los jugadores.

Primero se aplicará un Análisis de Componentes Principales a los datos y más tarde se buscarán perfiles de juego con K-medias.

En términos más sencillos, se reduce el número de variables contenidas en las estadísticas deportivas de forma que se pierda la menor cantidad de información posible y mediante modelado estadístico se segmenta en grupos de jugadores lo más homogéneos posibles, buscando perfiles de juego similares.

Cabe destacar que para que el modelo sea capaz de indentificar patrones solamente asociados al estilo de juego, evitando dividir en estrellas y jugadores promedio, se ha entrenado al modelo usando las siguientes estadísticas por minuto y que reflejan mejor el perfil del jugador:

Estas nuevas métricas son:

- **Proporciones de tiro:**

- FGA_per_MIN: Proporción de tiros de campo intentados por minuto.
- FG3A_rate: Proporción de tiros que son intentos de triple ($\frac{FG3A}{FGA}$).
- FTA_rate: Proporción de tiros libres generados por intento de tiro ($\frac{FTA}{FGA}$).

- **Distribución de rebotes:**

- OREB_rate: Proporción de rebotes que son ofensivos, respecto al total de rebotes capturados ($\frac{OREB}{OREB+DREB}$).

- **Estilo de juego:**

- AST_per_MIN: Asistencias por minuto.
- TOV_per_MIN: Pérdidas por minuto.
- AST_TOV_ratio: Relación asistencias/pérdidas, ajustada para evitar división por cero ($\frac{AST}{\max(TOV,1)}$).

- **Contribución defensiva:**

- STL_per_MIN: Robos por minuto.
- BLK_per_MIN: Tapones por minuto.

- **Uso ofensivo estimado:**

- Usage_rate: Tasa de uso ofensivo por minuto, basada en tiros de campo, tiros libres, asistencias y pérdidas:

$$\text{Usage_rate} = \frac{FGA + 0.44 \cdot FTA + AST + TOV}{MIN}$$

Se trata de una métrica avanzada usada en la NBA.

En nuestro caso se ha dividido a los jugadores en 5 grupos aunque podrían haber sido unos 6 o 7 perfectamente. La división obtenida es la siguiente:

Informe de Resultados

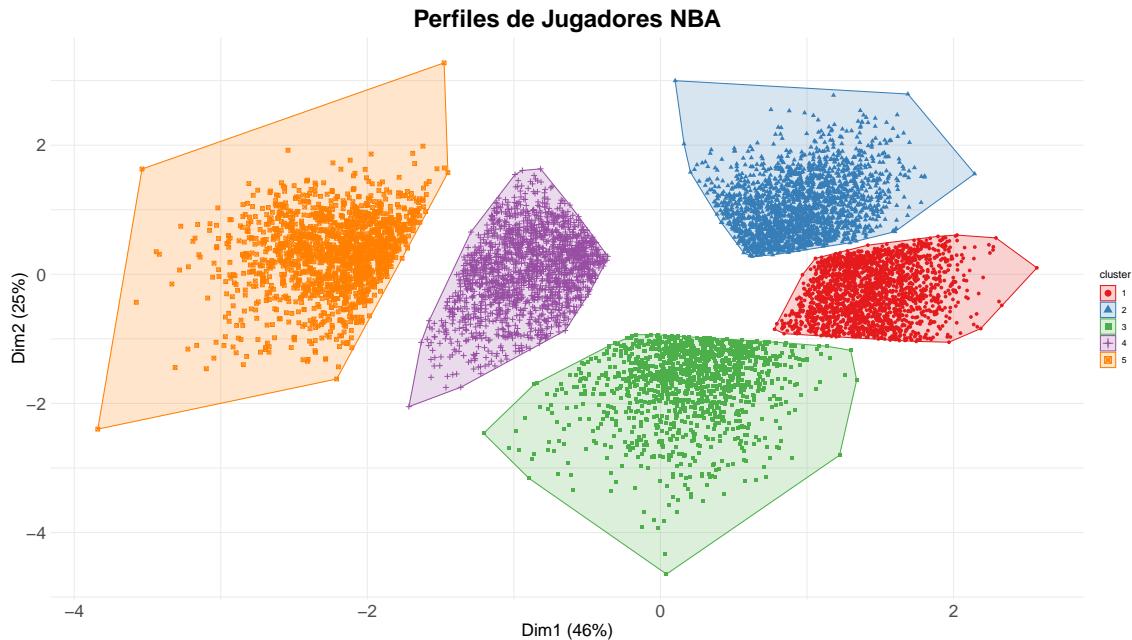


Figura 10

Las estadísticas de los diferentes grupos se pueden observar mediante gráficos de barras como se ha hecho en el análisis exploratorio:

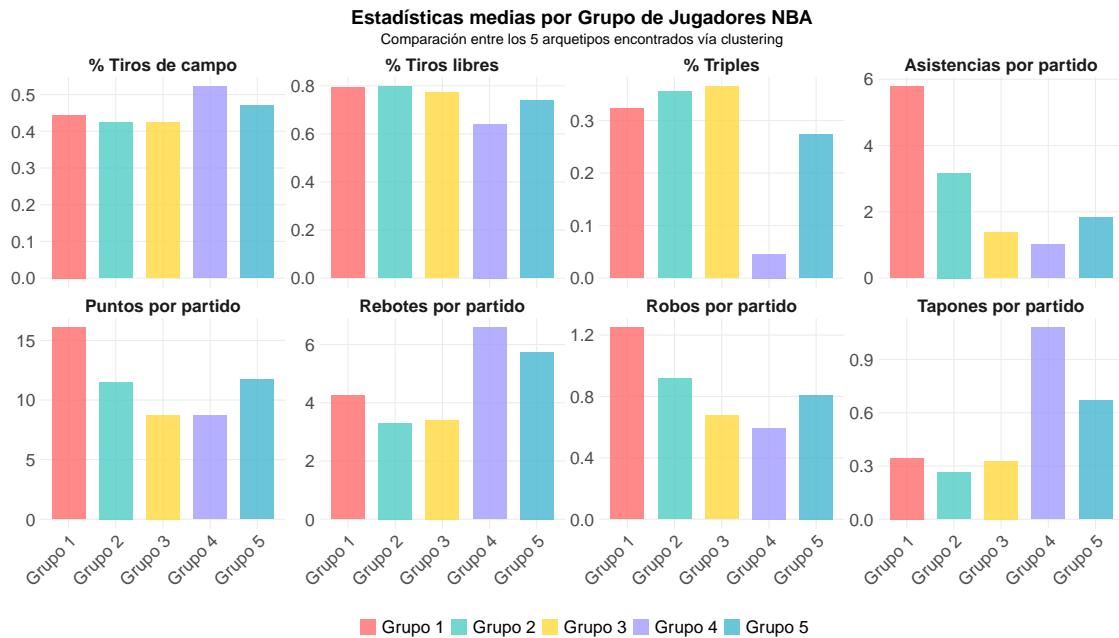


Figura 11

A continuación se comentan los aspectos generales e integrantes de los clústeres:

- **1. Creadores de juego. Exteriores y aleros híbridos:** Jugadores que dominan todo tipo de tiro, con mucha anotación por partido y líderes destacados en robos y asistencias. Ejemplos en el clúster: Jason Kidd, John Stockton, Steve Nash, Chris Paul, Allen Iverson, Dwyane Wade, Stephen Curry, Derrick Rose, Kobe Bryant, Scottie Pippen.

tie Pippen, LeBron James, Manu Ginóbili, Tracy McGrady, Nikola Jokić (sorprendente por ser un pívot pero encaja con su estilo), etc.

- **2 Bases y escoltas versátiles:** buenos tiradores desde el perímetro y con un conjunto de estadísticas equilibradas. Ejemplos: James Harden, Ray Allen, Gary Payton, Kemba Walker, Damian Lillard, Klay Thompson, Rudy Fernández, Michael Jordan, etc.
- **3. Tiradores de rol:** jugadores que destacan por su ayuda al equipo y sobretodo por su tiro de tres. Destacan Danny Green, Kevin Love, Dirk Nowitzki (en sus últimos años), JR Smith, etc. Se trata de jugadores que suelen jugar de 6, es decir, el primer recambio disponible en el banquillo, muy importante en la NBA moderna.
- **4. Big Men puros:** en este grupo predomina el pívot puro, defensivo, reboteador y con influencia cerca de canasta. Destaca por ser el grupo que promedia más tapones, rebotes y porcentaje de tiros de campo, debido a que los realizan muy cerca del aro. Su aportación en triples es nula. Ejemplos: Tim Duncan, Shaquille O'Neal, Yao Ming, Pau Gasol, Dikembe Mutombo, Dwight Howard, etc.
- **5. Defensa híbrida. Combinación de exteriores e interiores defensivos:** jugadores que destacan en varias facetas de juego, defensiva y anotadora, a diferencia de los jugadores anteriores, anotan triples ocasionalmente. Ejemplos: Vince Carter, Pau Gasol (de nuevo), Kevin Garnett, Nicola Jokić (de nuevo), etc.

Este modelo puede ir dirigido a equipos de *scouting* de las franquicias NBA para poder detectar ciertos perfiles de jugadores universitarios o fuera de la liga según las necesidades del equipo.

4 Estudio del All-Star

El *All-Star Game* de la NBA es un partido que se celebra todas las temporadas y que es disputado por los mejores jugadores de la misma. Aún siendo un evento de exhibición sin mucha competitividad, supone un gran reconocimiento para los mejores jugadores formar parte de él. Las casas de apuestas crean cuotas acerca de quién será elegido en cada temporada. Hay jugadores que son evidentes, pero las últimas elecciones pueden estar más disputadas. Por esto mismo, en esta sección se presenta un modelo que trata de predecir si un jugador será *All-Star* o no.

Primero se va a realizar una visualización que va a permitir ver si hay una diferencia real entre las estrellas y el resto de jugadores de la Liga². Se compararán:

- Estadísticas principales como puntos, rebotes, asistencias y pérdidas.
- Minutos jugados por partido. Las estrellas deberían tener más volumen de juego que los jugadores promedio, pues sus franquicias dependen de ellos.
- Una métrica avanzada de baloncesto llamada *True Shooting*. Se mide en porcentaje y considerando los tres tipos de tiro. Mide la eficiencia anotadora de un jugador. Su fórmula es:

$$TS \% = \frac{PTS}{2 \times (FGA + 0.44 \times FTA)} \times 100$$

El *True Shooting* es una métrica que normalmente se encuentra cercana al 50 %, aquellos que logran pasar la barrera del 60 % se pueden considerar de los mejores anotadores de la historia en cuanto a acierto en todas las disciplinas de tiro se refiere.

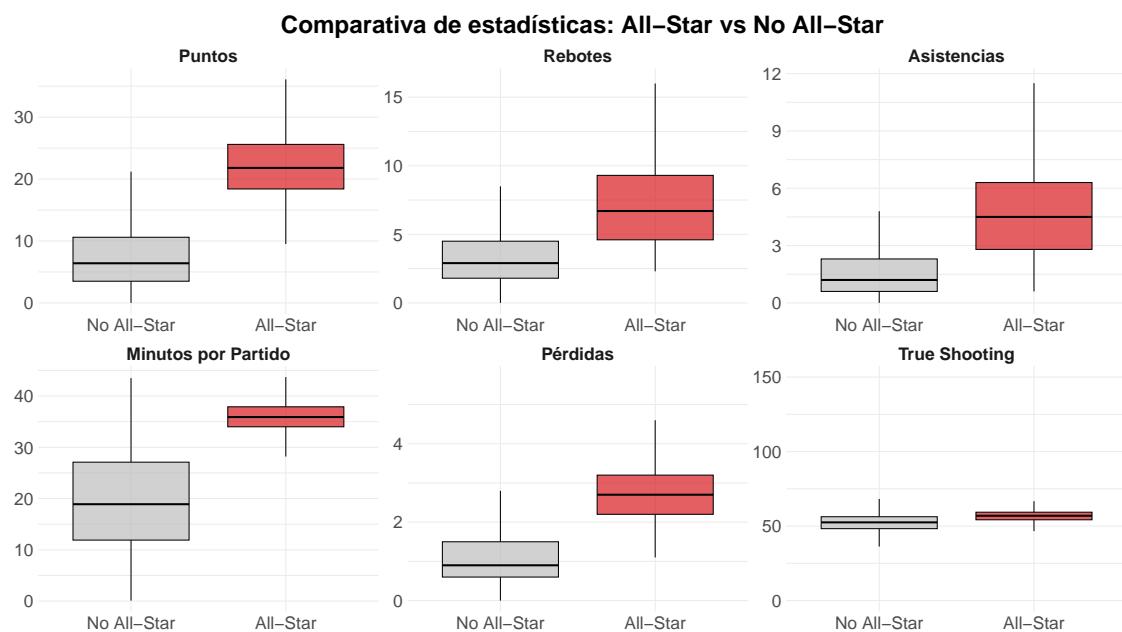


Figura 12

²Para evitar sesgos, en esta representación se han omitido jugadores de la temporada del 1998-99, en la cual no se disputó el All-Star

Informe de Resultados

Los diagramas de caja anteriores muestran que hay una gran diferencia entre jugadores *All-Star* y aquellos no elegidos. Curiosamente, las estrellas son los jugadores que más balones pierden. El *True Shooting* parece ser ligeramente más alto, pero en esta métrica esa diferencia significa mucho.

Aún con estas diferencias, hay algunos jugadores no elegidos *All-Star* que alcanzan a las estrellas en alguna que otra estadística, luego se ha de tener especial cuidado con ello.

Otro problema es que la cantidad de jugadores *All-Star* es mucho menor a la cantidad de jugadores no elegidos, que domina la muestra. Esto puede hacer que el modelo simplemente aprenda a predecir que un jugador no es *All-Star* en más casos de los que se desearía. Incluso con una alta tasa de acierto del modelo se deben acudir a otras métricas para poder evaluar su calidad debido a este aspecto.

En este caso se usará *Random Forest*, un algoritmo de aprendizaje automático supervisado que combina múltiples árboles de decisión para mejorar la precisión de las predicciones.

En la siguiente tabla se muestran los resultados al aplicar el modelo en un conjunto de datos distintos a los que han formado parte de su entrenamiento:

Referencia	No All-Star	All-Star
Predicción: No All-Star	2141	24
Predicción: All-Star	58	108

Cuadro 1: Predicciones frente a valores reales.

Si nos limitamos a la precisión del modelo, esta es del 96.48 %, lo cual está realmente bien. Sin embargo, la precisión esconde las predicciones realizadas para jugadores que realmente formaron parte del *All-Star*, ya que los jugadores no seleccionados tienen mucho más peso en el análisis. Por ello, se recurre a las siguientes métricas que son de mucha utilidad en estas situaciones:

- La **precisión**: es el cociente entre verdaderos positivos y el total de predicciones positivas: $\text{Precisión} = \frac{VP}{VP+FP}$
- El **recall**: mide la proporción de verdaderos positivos correctamente identificados: $\text{Recall} = \frac{VP}{VP+FN}$.
- El **F1-score**: es la media armónica entre precisión y recall: $F1 = 2 \cdot \frac{\text{Precisión}\cdot\text{Recall}}{\text{Precisión}+\text{Recall}}$. Ofrece un balance entre ambos cuando hay desequilibrio de clases.

En este caso se obtiene:

- Una **precisión** de la etiqueta positiva del 65 %. Es decir, de los jugadores predichos *All-Star*, el 65 % realmente lo son.
- Un **recall** del 81.81 %, lo cual indica el porcentaje de jugadores que han sido *All-Star* y que realmente han sido predichos como tal. Se trata de un muy buen valor ya que indica que el modelo construido no está discriminando a la etiqueta positiva por ser la minoritaria y a sus valores les está asignando la mayoritaria, sino que en 4 de cada 5 casos clasifica a los jugadores estrella como tal.

Informe de Resultados

Por tanto, se tiene un modelo que puede ser muy útil para evaluar que tan probable es que un jugador sea All-Star o no, lo cual puede ser muy útil para casas de apuestas o personas que se dedican al mundo de las apuestas deportivas.

5 Estudio del Draft

Primero se va a tratar de responder mediante una simple visualización a la cuestión de qué universidades son las más influyentes en el *draft*. Esto permite a los ojeadores de las franquicias poner especial atención en los equipos de estas instituciones en busca de talento joven.

El siguiente mapa muestra las universidades con mayor volumen de jugadores incorporados a la NBA y además mediante el color se indica el *pick* medio de sus elecciones en el *draft*. El mapa incluye datos de las últimas 4 décadas, desde los 80 hasta la década del 2010.



Figura 13: **Nota:** el mapa está en formato PDF, acercar para verlo mejor

Las universidades de la parte Este de los Estados Unidos dominan el *draft* de la NBA sobre zonas de interior y de la costa Oeste.

Las universidades que más talento aportan son Duke, Kansas, Kentucky, Arizona y UCLA, entre otras. Algunas como Nebraska o Virginia han aportado un número muy pequeño de jugadores a la liga durante este tiempo.

Las universidades con los *picks* más altos en promedio son Texas, Duke, Kentucky, Utah y UNLV. San Diego State y Pepperdine aportan algunos jugadores a la liga, pero en *picks* tardíos de la segunda ronda del *draft* en su mayoría.

A continuación, se enfrenta el orden de elección en el *draft* frente a los años que posteriormente los jugadores pasaron en la NBA. Con este fin, se pretende ver si realmente *picks* más altos suelen tener más éxito en la liga. Los datos mostrados corresponde a jugadores elegidos en los draft de 2001, 2002 y 2003.

Informe de Resultados

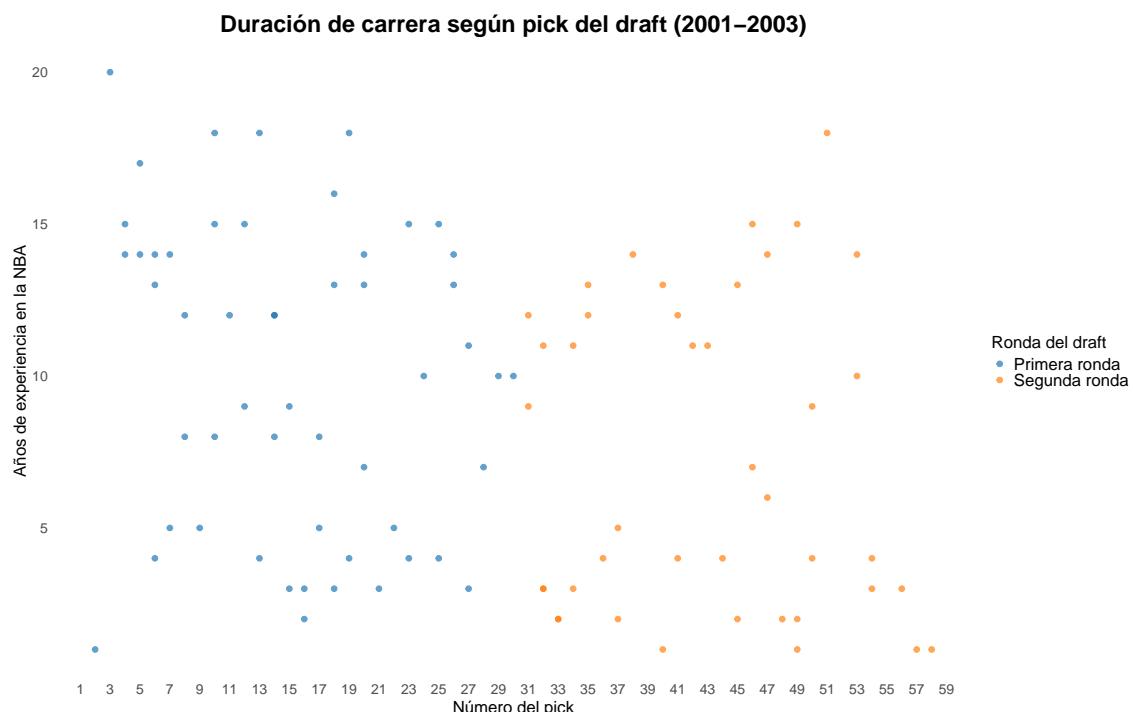


Figura 14

Como se puede apreciar, no hay una diferencia abismal en cuanto a sobrepasar las 5 primeras temporadas, pues muchos jugadores elegidos en la segunda ronda han jugado incluso más de 10 años en la NBA. Esto quiere decir que en el *draft* es posible encontrar joyas ocultas que pasan desapercibidas en la primera ronda y que luego acaban siendo jugadores consolidados de la NBA³.

Distribución por clase académica en NCAA

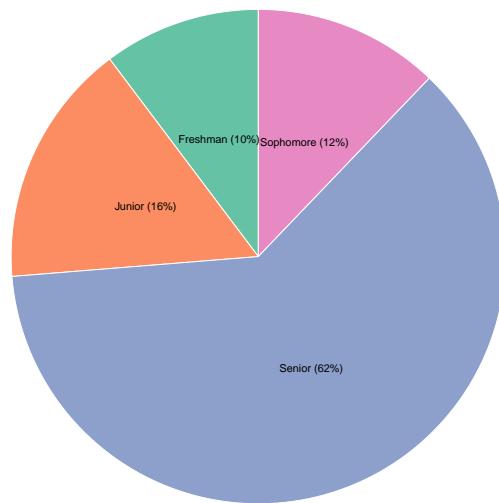


Figura 15: Distribución por año académico de los jugadores seleccionados en el *draft*.

³En la gráfica no se aprecian jugadores elegidos en primer lugar porque LeBron James y Kwame Brown entraron al finalizar su educación secundaria y Yao Ming vino de China.

La gráfica de sectores anterior muestra que la mayoría de universitarios que acceden a la NBA lo hacen en su cuarto año al graduarse (*Senior*). Lo inusual es que accedan en primer, segundo o tercer año.

Es lógico preguntarse si los jugadores que entran a la liga como *Senior* están más preparados y tienen carreras más duraderas que los que lo hacen al principio de su etapa universitaria. Aunque parezca contraintuitivo, ocurre lo contrario.

Según los datos, los *Junior* (estudiantes de tercer año) permanecen de media unos 8.3 años en la NBA, mientras que los de primer y segundo año lo hacen unas 7 temporadas, quedando los *Senior* con una media de experiencia de 6 años y medio. Esto puede deberse a que jugadores *Senior* no son elegidos durante sus estudios y prueban suerte en su última oportunidad de acceder a la liga. Por otro lado, los estudiantes que aún no han acabado sus estudios son elegidos porque realmente destacan entre los jugadores disponibles.

En cuanto a modelado estadístico se ha usado *XGBoost*, esta vez como modelo de clasificación, para poder predecir si un jugador tendrá éxito en la liga o no, dándose este cuando su experiencia total llega a ser, al menos, de 5 años. La siguiente tabla muestra el rendimiento del modelo a la hora de clasificar:

Referencia	No Éxito	Éxito
Predicción: No Éxito	57	38
Predicción: Éxito	47	159

Cuadro 2: Matriz de confusión del modelo: predicciones frente a valores reales.

Se obtiene una precisión del 71.76 %. En este caso la etiqueta positiva es tener éxito en la NBA. Se obtienen así los siguientes valores de rendimiento:

- Recall de 80.71 %.
- Precisión de la etiqueta positiva de 77.18 %
- Un **F1-Score** de 0.79

Lo cual quiere decir que el modelo es aceptable, pues solamente se ha entrenado con estadísticas de juego universitario y generaliza bastante bien normalmente. El pick en el que fue elegido en jugador no ha sido usado, pues no tendría sentido para poder hacer de este modelo una aplicación útil. La ronda sí se ha tomado como variable porque antes del *draft* siempre se intuye en qué ronda va a ser elegido dicho jugador.

Por último se muestra la gráfica con la importancia de las variables predictoras involucradas:

Informe de Resultados

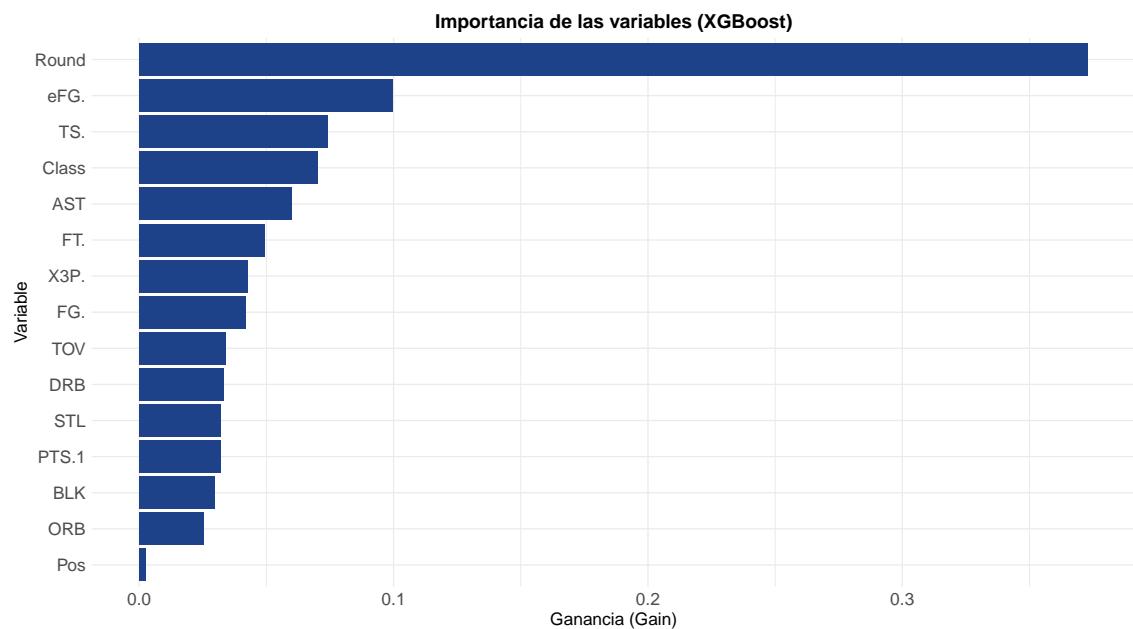


Figura 16

Los pronósticos de ronda de elección son muy importantes, al igual que métricas como el *True Shooting*, que representa de muy buena manera la eficiencia anotadora de un jugador. La clase académica también cobra bastante importancia a la hora de predecir el éxito, mientras que la posición de juego apenas influye en las decisiones del modelo.

6 Conclusión

Este estudio ha permitido abordar, desde una perspectiva cuantitativa y basada en datos reales, diferentes aspectos clave del ecosistema NBA. A través de un proceso exhaustivo de extracción, limpieza, análisis y modelado, se ha construido un conjunto de datos completo que integra estadísticas de juego, información personal, historial universitario, salarios y participación en eventos como el *All-Star*. Este enfoque ha facilitado un análisis integral del valor y potencial de los jugadores a lo largo de más de 25 temporadas.

Uno de los principales hallazgos ha sido la complejidad que encierra el salario de un jugador NBA, el cual depende no solo del rendimiento en pista, sino también de factores contractuales, marketing, trayectoria y contexto de la franquicia. Pese a ello, los modelos desarrollados han sido capaces de ofrecer predicciones razonables del salario únicamente a partir de estadísticas deportivas, mostrando que existe una base objetiva sobre la que fundamentar el valor económico de un jugador.

Asimismo, se ha demostrado que es posible anticipar con cierto grado de acierto qué jugadores tienen más probabilidades de alcanzar el estatus de *All-Star* o de consolidarse profesionalmente tras su paso por la NCAA. Esto pone de manifiesto el valor de la analítica en la toma de decisiones estratégicas, tanto a nivel de *scouting* como de gestión de plantilla.

En definitiva, este trabajo evidencia cómo el uso riguroso de herramientas estadísticas puede aportar claridad en un entorno marcado por la incertidumbre, el talento y la competencia. La NBA, al ser una liga con una cultura de datos tan asentada, se presenta como un escenario idóneo para este tipo de análisis, con potencial de aplicación tanto en el ámbito académico como en el profesional.

Referencias

- [1] National Basketball Association. *Sitio web oficial de la NBA y recursos multimedia.* Disponible en: <https://www.nba.com>
- [2] NBA API. *An API client package for NBA.com data.* Disponible en: https://github.com/swar/nba_api
- [3] HoopsHype. *NBA player salaries, rumors and stats.* Disponible en: <https://hoopshype.com>
- [4] Wikipedia. *2024 NBA All-Star Game.* Disponible en: https://en.wikipedia.org/wiki/2024_NBA_All-Star_Game
- [5] Stathead Basketball. *Advanced basketball statistics and research tools.* Disponible en: <https://www.stathead.com/basketball/>
- [6] D. Wang. *Predictive Modeling of NBA Player Salaries Using Machine Learning.* 2023. Disponible en: <https://medium.com/@dwang22>
- [7] The R Graph Gallery. *Inspiration and code for R charts.* Disponible en: <https://www.r-graph-gallery.com>