Imagen que contiene Logotipo

El contenido generado por IA puede ser incorrecto.

Contenido

[Resumen 3](#_heading=h.cuirqopg5ony)

[Objetivo del proyecto 4](#_heading=h.re3i8u8oorq)

[Datos y metodología 4](#_heading=h.280lbvy03koa)

[Origen de los datos 4](#_heading=h.ykg5pm43a6yy)

[Archivos seleccionados 5](#_heading=h.lccshvvvonwo)

[Limpieza y transformación 5](#_heading=h.91pb7b6boklo)

[Tratamiento de valores faltantes 5](#_heading=h.97nkynbadaei)

[Análisis estadístico 5](#_heading=h.9mzlq5lr14jd)

[Visualizaciones 6](#_heading=h.umsw7x56jl08)

[Análisis exploratorio de datos (EDA) e insights obtenidos 7](#_heading=h.pz3xvi3k4fs3)

[Contexto histórico de la liga 7](#_heading=h.n3793m6awmtq)

[Evolución de la cantidad de jugadores por temporada 8](#_heading=h.i5s4hixba2l3)

[Distribución general del rendimiento por jugador y temporada 9](#_heading=h.uocde349q0lp)

[Top jugadores por eficiencia (Net Rating) 10](#_heading=h.9iaynywm1y14)

[Tendencia Histórica de eficiencia promedio 11](#_heading=h.xtlo8l2i57qw)

[Correlación entre métricas de rendimiento 12](#_heading=h.adord3khpvac)

[Análisis de rendimiento colectivo 13](#_heading=h.uyzc5l7brw1t)

[Distribución del puntaje global por temporada 13](#_heading=h.kc7t7uz3m5ge)

[Evolución del percentil 10 (vara) por equipo 14](#_heading=h.jvth8w1qesnk)

[Indicadores y métricas creadas 15](#_heading=h.g5m20rv0kf4)

[Rendimiento Más Óptimo (RMO) 15](#_heading=h.adzd5026c39c)

[Clasificación de Jugadores según su rendimiento vs. RMO 15](#_heading=h.az2y2gq8hyz8)

[Distribución de eficiencia con umbral RMO 16](#_heading=h.veynmb6a04hx)

[Top 10 Jugadores con rendimiento superior al RMO 17](#_heading=h.rcjrufb4phbx)

[Revisión de trayectorias 18](#_heading=h.eejz10pifao5)

[Modelo Predictivo de Rendimiento Global 18](#_heading=h.udzirnkznjd2)

[Métricas de Evaluación del Modelo 18](#_heading=h.fq80wjyr644r)

[Segmentación de Jugadores mediante Clustering (K-Means) 19](#_heading=h.jxjzqr5jhjq4)

[K-Means es un algoritmo de machine learning no supervisado que: 19](#_heading=h.u1dn1v61nz7t)

[Análisis de Clústeres 20](#_heading=h.j9a6bmyjv7cu)

[Conclusiones a partir del Clustering: 21](#_heading=h.henap9y6ms6g)

[Aplicación estratégica del clustering 21](#_heading=h.6cgcph6uwydx)

[Métrica dominante por clúster 21](#_heading=h.b9kd6fmdrr0m)

[Recomendaciones estratégicas 22](#_heading=h.vdm41xy8jdi3)

[Priorización del Talento Ofensivo (Adaptación a la Tendencia) 22](#_heading=h.nrr1p9yawck9)

[Gestión de Riesgo en la Rotación (El "Factor 12.79%") 23](#_heading=h.jzfkywpo0lxa)

[Scouting basado en Clustering y Perfiles Complementarios 23](#_heading=h.pkt3rcm88b96)

[Benchmarking de Franquicias Exitosas 24](#_heading=h.1rjksuogcehq)

[Valoración Objetiva vs. Subjetiva en Renovaciones 24](#_heading=h.3qmcy86vqpxv)

[Inversión en Eficiencia Ofensiva sobre Volumen 25](#_heading=h.a2eznycat0ug)

[Estrategia Diferenciada para Talento Subvalorado 25](#_heading=h.om1c62a8ph2h)

[Monitoreo del "Vara" como Indicador de Salud Organizacional 25](#_heading=h.yoet3lcj2ap0)

[Adaptación a la Era Post-Pandemia 25](#_heading=h.snc630h4vpnr)

[Navegación del Dashboard 26](#_heading=h.irad69lzp3c8)

[Rendimiento Individual del Equipo 26](#_heading=h.3rqs5osgnrkq)

[Panorama por equipo en la temporada 27](#_heading=h.ysuhwow361dx)

[Evolución Histórica y Temporal 28](#_heading=h.rbh4pwmtlzo7)

[Limitaciones del análisis 30](#_heading=h.v4uxhvqj4p88)

[Alcance temporal de los datos 30](#_heading=h.kcrgr7t4kb50)

[Contexto de Equipo Limitado 30](#_heading=h.8c3v4z2jna8)

[Normalización por Minutos Jugados 31](#_heading=h.yjqr5ckw9u0p)

[Datos Faltantes en temporadas tempranas 31](#_heading=h.490j1m3vei6s)

[Variables no cuantificables 31](#_heading=h.aexnjegcsjck)

[Granularidad de Posiciones 32](#_heading=h.vt9kpuk5bei8)

[Anomalías y errores en datos originales 32](#_heading=h.pao4d7dh16qi)

[Generalización a otras ligas 32](#_heading=h.m56te3dgsxxu)

[Conclusiones 33](#_heading=h.s7aonziz7bh0)

[Implicaciones estratégicas para la industria 34](#_heading=h.a4fp3uya30r0)

[Para gerencias deportivas: 34](#_heading=h.xwwsq7no4dqn)

[Para entrenadores y cuerpo técnico: 34](#_heading=h.zc6t8lr98iop)

[Para analistas y científicos de datos: 34](#_heading=h.dt5ja3rtyv8c)

[Visión de futuro: el Camino hacia decisiones verdaderamente inteligentes 34](#_heading=h.19w9vi1ij7j9)

**Decisiones inteligentes NBA: Análisis de talento y rendimiento**

# Resumen

El presente informe desarrolla un análisis integral del rendimiento de jugadores y equipos pertenecientes a la National Basketball Association (NBA), la principal liga profesional de baloncesto en el mundo. A partir de un conjunto de datos estadísticos que abarca 26 temporadas (1996-2022), se realizó un proceso completo de limpieza, transformación, análisis exploratorio y modelado predictivo con el objetivo de identificar patrones de desempeño individual y colectivo.

Los resultados muestran diferencias importantes en la contribución ofensiva y defensiva de los jugadores, niveles de consistencia por partido, eficiencia de tiro, volumen de posesiones y otras métricas que influyen directamente en el resultado de los encuentros. También se evidenciaron relaciones entre estadísticas clave—como puntos, rebotes, asistencias, efectividad, rating ofensivo y rating defensivo—que permiten comprender los factores determinantes para el desempeño competitivo de los equipos.

El análisis incorpora técnicas avanzadas de machine learning, incluyendo modelos predictivos con 86% de precisión (R²=0.8613) y segmentación mediante clustering K-Means que identifica cuatro perfiles diferenciados de jugadores. Se establecieron umbrales objetivos de rendimiento (RMO - Rendimiento Más Óptimo) que permiten clasificar jugadores en categorías estratégicas.

Los hallazgos obtenidos proporcionan una base sólida para apoyar decisiones estratégicas relacionadas con renovación de contratos, cambios de rol, scouting, transferencias o reestructuración de los equipos. En conjunto, el análisis ofrece una visión objetiva y cuantitativa del rendimiento dentro de un entorno altamente competitivo como la NBA.

# Objetivo del proyecto

Desarrollar un informe analítico que permita visualizar patrones de desempeño individual y colectivo, con el fin de tomar decisiones futuras como renovación, retiro o reestructuración de equipos.

# Datos y metodología

## Origen de los datos

Los datos fueron recopilados desde la plataforma Kaggle, una fuente reconocida de datasets abiertos para análisis exploratorio y modelado. El repositorio seleccionado contenía 16 archivos en formato CSV, con información estadística detallada de partidos, jugadores, equipos y métricas avanzadas de rendimiento que abarca desde la temporada 1996-97 hasta la 2022-23. Sin embargo, no todos los archivos eran necesarios para cumplir el objetivo analítico del proyecto. Por ello, se realizó un proceso de revisión preliminar del contenido de cada archivo, verificando:

* La calidad de los datos
* La estructura y completitud
* La relevancia de las variables para el análisis de rendimiento
* La compatibilidad entre archivos
* El volumen de registros disponibles

Como resultado de esta revisión, se seleccionaron 5 archivos principales, considerados los más pertinentes para el análisis. Estos archivos contienen información clave como estadísticas individuales por jugador, estadísticas por equipo, resultados de partidos y métricas avanzadas de desempeño.

* Universo de análisis:
* Total de jugadores analizados: 13,997
* Temporadas cubiertas: 26 (1996-97 a 2022-23)
* Equipos históricos y activos: 36 franquicias
* Selección de archivos relevantes

Los criterios para seleccionar los 5 archivos principales fueron:

* Inclusión de métricas directamente asociadas al rendimiento ofensivo y defensivo
* Disponibilidad de identificadores consistentes para unir información entre archivos
* Volumen suficiente de registros para generar conclusiones confiables
* Menor presencia de valores faltantes o inconsistencias
* Relevancia respecto al objetivo del proyecto (evaluación de desempeño individual y colectivo)

## Archivos seleccionados

A continuación, se mencionan los archivos seleccionados:

* *all\_seasons.csv* - Estadísticas consolidadas por jugador y temporada
* *player.csv* - Información demográfica, física y biográfica de jugadores
* *line\_score.csv* - Resultados detallados por partido
* *game.csv* - Información general de encuentros
* *team.csv* - Estadísticas agregadas por equipo y temporada

Esta selección permitió concentrar el análisis en datos limpios, relevantes y comparables, garantizando una base sólida para las siguientes etapas.

## Limpieza y transformación

La fase de limpieza y transformación de los datos se realizó íntegramente en Python, utilizando librerías especializadas como Pandas y NumPy. Todos los archivos seleccionados del repositorio de Kaggle se trabajaron en formato CSV, lo cual facilitó su lectura, manipulación y posterior integración dentro del notebook del proyecto. Para optimizar el proceso y distribuir responsabilidades, cada integrante del equipo asumió la limpieza, revisión y análisis preliminar de uno de los archivos seleccionados. Esta dinámica permitió una comprensión profunda de la estructura, calidad y características específicas de cada dataset antes de integrarlos en el análisis colectivo.

El proceso general de limpieza incluyó:

* Carga individual de los archivos CSV en Python
* Revisión de la estructura de cada archivo (dimensiones, variables, tipos de datos)
* Corrección de tipos de datos (conversión a numéricos, fechas, strings estandarizados)
* Eliminación de registros duplicados identificados mediante claves compuestas
* Identificación y tratamiento de valores nulos, según el comportamiento de cada archivo
* Normalización de nombres de columnas para facilitar uniones entre archivos
* Estandarización de códigos de equipos (considerando relocalizaciones históricas)
* Creación de columnas adicionales, como métricas personalizadas de rendimiento

## Tratamiento de valores faltantes

* Variables numéricas: Imputación por mediana cuando representaba <5% del total
* Variables categóricas: Análisis caso por caso según contexto
* Registros incompletos críticos: Eliminados si afectan métricas clave (representaron <2% del total)

## Análisis estadístico

Se aplicaron técnicas estadísticas descriptivas e inferenciales para comprender la distribución y relaciones entre variables:

* Estadística descriptiva:
* Medidas de tendencia central (media, mediana, moda)
* Medidas de dispersión (desviación estándar, rango intercuartílico)
* Análisis de percentiles para establecer umbrales de rendimiento
* Estadística inferencial:
* Análisis de correlaciones de Pearson entre métricas clave
* Pruebas de normalidad para validar distribuciones
* Identificación de valores atípicos mediante método IQR
* Técnicas avanzadas:
* Estandarización z-score para crear métricas comparables
* Análisis de componentes principales (considerado para reducción dimensional)
* Clustering K-Means para segmentación de perfiles
* Regresión lineal múltiple para modelado predictivo}

## Visualizaciones

Se desarrolló un ecosistema completo de visualizaciones dividido en dos categorías:

**Categoría 1:** Análisis Exploratorio (EDA) en Python:

* Histogramas con curvas de densidad (KDE)
* Gráficos de dispersión (scatter plots)
* Mapas de calor de correlaciones
* Boxplots temporales
* Gráficos de barras comparativos
* Gráficos de líneas para tendencias históricas

**Categoría 2:** Dashboards Interactivos:

* Panel de "Panorama por Equipo en la Temporada"
* Panel de "Rendimiento Individual del Equipo"
* Panel de "Evolución Histórica y Temporal"
* Todas las visualizaciones fueron diseñadas siguiendo principios de claridad, accesibilidad visual y capacidad de contar historias con datos (data storytelling).
* Análisis exploratorio de datos (eda) e insights obtenidos

# Análisis exploratorio de datos (EDA) e insights obtenidos

## Contexto histórico de la liga

****

*Gráfica 1. Cantidad de partidos por temporadas*

Este gráfico permite visualizar la evolución del número de partidos por temporada desde 1996 hasta 2022.

**Hallazgos clave:**

* Caídas notables en 1998, 2011 y especialmente en 2012, que presenta el menor número de partidos. Esto podría estar relacionado con huelgas, cambios en el calendario o eventos externos como crisis económicas o sanitarias.
* Recuperación progresiva después de 2012, alcanzando niveles similares a los de las décadas anteriores.
* Estabilidad general en la mayoría de las temporadas, lo que sugiere una estructura de liga relativamente constante**.**

**Posibles causas de estos hallazgos**

|  |  |  |  |
| --- | --- | --- | --- |
| **Año (Aprox)** | **Evento** | **Causa** | **Efecto en la gráfica** |
| **1998-99** | Lockout | Disputa laboral | Caída fuerte (50 juegos/equipo). |
| **2011-12** | Lockout | Disputa laboral | Caída media (66 juegos/equipo). |
| **2012 (ID)** | *Error de Datos* | Probable error en el dataset | *Anomalía:* La caída es demasiado profunda para ser real; la temporada 12-13 fue normal. |
| **2019-21** | COVID-19 | Crisis Sanitaria | Caída y fluctuación (Suspensión y 72 juegos/equipo). |

*Tabla 1. Posibles causas del comportamiento de datos asociados a partidos por temporada*

Este contexto es fundamental para interpretar correctamente las métricas de rendimiento. Las temporadas afectadas por lockouts o la pandemia deben analizarse considerando el calendario reducido, que puede sesgar promedios y totales acumulados.

## Evolución de la cantidad de jugadores por temporada

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

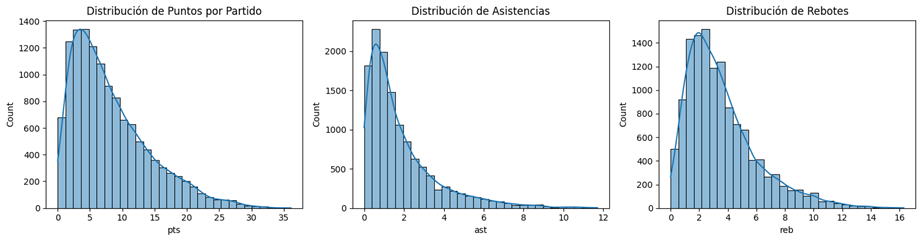
*Gráfica 2. cantidad de jugadores por temporada*

Este gráfico muestra la cantidad total de registros de jugadores por año de temporada, desde 1996 hasta 2019. Cada barra representa el volumen de participación o cobertura de jugadores en la base de datos. En esta gráfica se observa:

* Tendencia creciente: A lo largo de las temporadas, se observa un aumento sostenido en el número de registros. Esto sugiere una expansión progresiva en la cobertura de datos, mayor rotación de jugadores, o una mejora en la recopilación histórica
* Pico reciente: A partir de 2010, el crecimiento se acelera, alcanzando su punto más alto entre 2017 y 2019, donde se superan los 500 registros por temporada
* Consistencia estructural: Las primeras temporadas presentan menor volumen, lo que puede deberse a limitaciones en la disponibilidad de datos históricos o menor número de jugadores activos registrados

Este gráfico permite interpretar los análisis de rendimiento. Un mayor número de registros por temporada puede influir en la dispersión de métricas como global\_score, vara y elite, y permite evaluar la evolución de la profundidad competitiva en la liga. Además, refuerza la solidez del dataset utilizado para el estudio.

## Distribución general del rendimiento por jugador y temporada

***Gráfica***  *3. Distribución general del rendimiento por jugador y temporada*

A partir de estos datos se pudo evidenciar cómo se distribuyen las principales métricas de rendimiento (puntos, asistencias y rebotes) por partido a lo largo de todas las temporadas registradas. Para esta parte se tiene: tres histogramas con curvas de densidad (KDE) que muestran la frecuencia de valores para cada métrica:

* Puntos por partido: La mayoría de los jugadores anotan entre 0 y 10 puntos por partido, con un pico alrededor de los 5 puntos. Esto refleja que solo una minoría alcanza cifras altas de anotación de forma constante.
* Asistencias por partido**:** Predominan los valores bajos, con la mayoría de los jugadores registrando entre 0 y 3 asistencias por partido. El pico se encuentra cerca de 1 asistencia, lo que indica que los roles de creación de juego están concentrados en unos pocos jugadores.
* Rebotes por partido: La distribución también está sesgada hacia la derecha, con un pico cerca de los 2 rebotes. Esto sugiere que los jugadores con alto volumen de rebotes son menos frecuentes.

Estas distribuciones reflejan la especialización de roles en el baloncesto profesional. La mayoría de los jugadores tienen rendimientos modestos en cada métrica, mientras que los valores extremos (jugadores con alto rendimiento) son menos comunes pero cruciales para el desempeño del equipo.

## Top jugadores por eficiencia (Net Rating)

Gráfico, Gráfico de embudo

El contenido generado por IA puede ser incorrecto.*Gráfica 4. Top jugadores por eficiencia*

Se hizo necesario identificar a los jugadores con mejor rendimiento promedio según el indicador de eficiencia (Net Rating), calculado a partir de sus estadísticas por partido (puntos, asistencias, rebotes). A partir de este análisis se tiene que:

* Ahmad Caver lidera el ranking con la eficiencia promedio más alta, seguido por Jerrelle Benimon y DeJon Jarreau
* Todos los jugadores del top presentan valores significativamente superiores al promedio general
* La métrica Net Rating permite comparar de forma integrada el aporte ofensivo y defensivo de cada jugador

Por lo anterior:

* Estos jugadores podrían considerarse referentes de eficiencia en sus respectivas temporadas
* El hecho de que algunos nombres no sean ampliamente conocidos sugiere que hay talento emergente o subvalorado que merece atención
* Útiles para análisis de scouting y decisiones de contratación

## Tendencia Histórica de eficiencia promedio

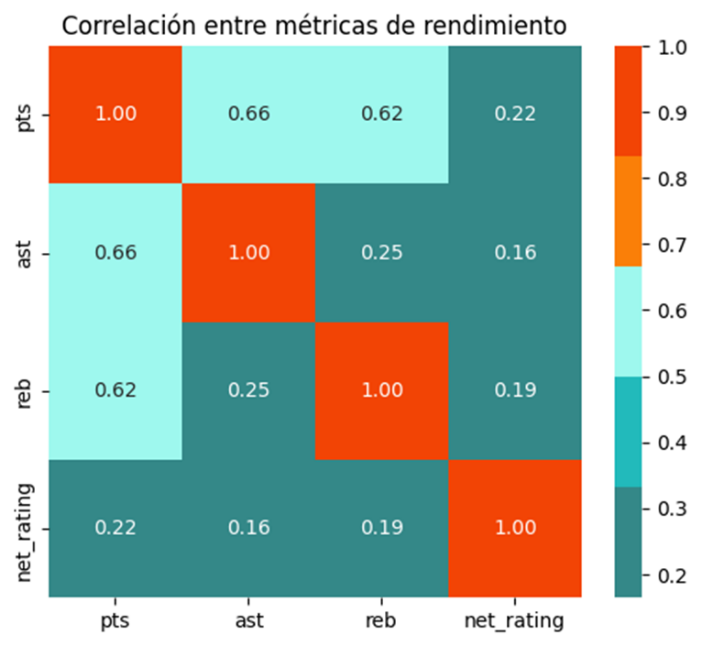
**Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.***Gráfica 5. Tendencia Histórica de eficiencia promedio*

Se consideró necesario observar cómo ha evolucionado la eficiencia promedio (Net Rating) de los jugadores a lo largo de las temporadas, desde 1996-97 hasta 2022-23. A partir de esto se identificó:

* Variabilidad significativa en la eficiencia promedio entre temporadas
* Caída marcada en la temporada 2021-22, posiblemente relacionada con factores externos como lesiones, cambios en el calendario o efectos post-pandemia
* Recuperación inmediata en 2022-23, indicando normalización del rendimiento
* La eficiencia promedio se ha mantenido en valores negativos, lo cual es esperable en métricas relativas que consideran el diferencial entre ofensiva y defensiva

## Correlación entre métricas de rendimiento



*Gráfica 6. Correlación entre métricas de rendimiento*

En este análisis es fundamental saber cómo se relacionan entre sí las principales métricas de rendimiento por partido: puntos (pts), asistencias (ast), rebotes (reb) y eficiencia (net\_rating). A partir de esto, se evidencia:

* Puntos y asistencias: Correlación moderada de 0.66 - los jugadores que anotan también tienden a generar juego
* Puntos y rebotes: Correlación de 0.62 - jugadores con alto volumen ofensivo también contribuyen en defensa o juego físico
* Asistencias y rebotes: Correlación baja (0.25) - refleja especialización de roles: los armadores no suelen ser los principales reboteadores
* Net Rating: Correlaciones débiles con las demás métricas (máximo 0.22 con puntos) - la eficiencia es una métrica compleja influenciada por múltiples factores

Con esta información se pudo entender cómo se complementan los roles dentro del equipo. La baja correlación entre net\_rating y otras métricas sugiere que la eficiencia debe analizarse en conjunto con contexto de equipo, minutos jugados y situaciones de juego. Así mismo, permitió identificar qué métricas pueden ser redundantes o independientes

## Análisis de rendimiento colectivo

A continuación, se presenta el análisis de rendimiento colectivo:

## Distribución del puntaje global por temporada

Gráfico, Gráfico de cajas y bigotes

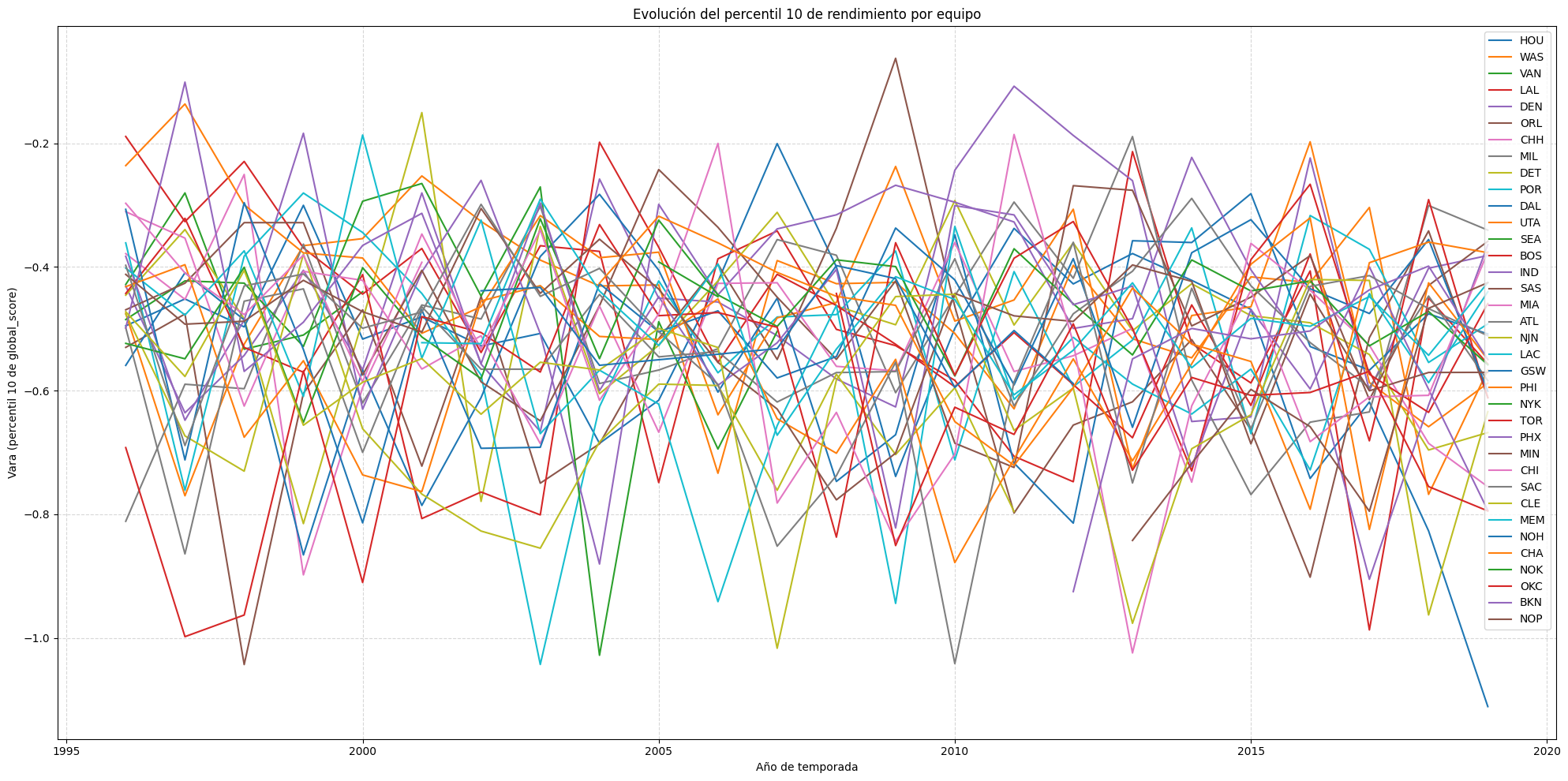
El contenido generado por IA puede ser incorrecto.*Ilustración 7. Distribución del puntaje global por temporada*

Con este análisis se puede evidenciar cómo varía el rendimiento relativo de los jugadores en cada temporada, desde 1996 hasta 2019, utilizando el puntaje global estandarizado (global\_score). Cada caja representa la dispersión de los puntajes dentro de una temporada específica. Se tiene que:

* Mediana estable: La línea central de cada caja indica que el rendimiento promedio por temporada se mantiene relativamente constante, lo que sugiere estabilidad en el nivel competitivo general
* Dispersión controlada: La mayoría de las temporadas presentan cajas de altura similar, reflejando una distribución homogénea del rendimiento entre jugadores
* Outliers consistentes: Los puntos fuera de las cajas representan jugadores con rendimiento excepcionalmente alto o bajo. Su presencia constante indica que cada temporada tiene figuras destacadas y desempeños fuera de norma

Comparar temporadas en términos de consistencia y variabilidad del rendimiento es útil para detectar años con mayor concentración de talento, cambios en el estilo de juego, o evolución en la eficiencia de los jugadores, reforzando la validez del puntaje global como métrica comparativa entre temporadas.

## Evolución del percentil 10 (vara) por equipo



*Gráfica 8. Evolución del percentil 10*

La vara representa el valor que marca el límite inferior del 10% de rendimiento (global\_score) dentro de cada equipo por temporada. Refleja el desempeño de los jugadores menos destacados. Esto significa:

* Tendencias ascendentes: Equipos cuya línea de vara sube con el tiempo muestran una mejora en la profundidad de su plantilla. Indica mejor reclutamiento, desarrollo de talento o estabilidad interna.
* Tendencias descendentes o fluctuantes: Caídas o variaciones abruptas pueden señalar temporadas de reconstrucción, lesiones, rotación excesiva o bajo rendimiento colectivo.
* Comparación entre franquicias: Algunas franquicias como SAS, BOS o MIA muestran trayectorias más estables, mientras que otras como VAN, SEA o NOK presentan discontinuidades relacionadas con cambios estructurales (reubicaciones, fusiones).

Con este análisis se pudo evaluar la solidez interna de cada equipo más allá de sus estrellas. Una vara alta indica que incluso los jugadores menos destacados tienen buen rendimiento, mientras que una vara baja puede reflejar debilidad estructural.

# Indicadores y métricas creadas

## Rendimiento Más Óptimo (RMO)

Esta métrica permite definir valores de referencia para identificar jugadores con rendimiento destacado en cada métrica, utilizando el percentil 75 como criterio (es decir, el 25% superior de todos los registros).

**Valores de RMO calculados:**

| **Métrica** | **RMO (Percentil 75)** | **Interpretación** |
| --- | --- | --- |
| Puntos (pts) | 11.5 | Cuartil superior de anotadores |
| Asistencias (ast) | 2.4 | Perfil destacado en generación de juego |
| Rebotes (reb) | 4.7 | Jugador fuerte en recuperación de balón |
| Eficiencia (net\_rating) | 3.2 | Impacto positivo sostenido en el equipo |

*Tabla 2. Valores de RMO calculados*

Esto permite:

* Estos umbrales pueden usarse para filtrar jugadores de alto rendimiento
* Construir perfiles tipo o establecer criterios de selección para scouting
* Sirven como base para segmentar la base de datos en grupos de rendimiento (élite, promedio, bajo impacto)

## Clasificación de Jugadores según su rendimiento vs. RMO

Con este cálculo se puede identificar y segmentar a los jugadores en tres categorías de rendimiento, comparando sus estadísticas individuales con los umbrales definidos por el percentil 75 (RMO).

**Criterios de clasificación:**

* **Alto rendimiento:** Jugadores que superan el RMO tanto en puntos como en eficiencia (net\_rating). Representan el grupo élite.
* **Rendimiento aceptable:** Jugadores que no alcanzan el RMO completo, pero tienen una eficiencia igual o superior al 85% del umbral. Son perfiles funcionales con impacto positivo.
* **Bajo rendimiento:** Jugadores que no cumplen con ninguno de los criterios anteriores. Representan la mayoría, con rendimiento limitado o especializado.

Con esto se pudo evidenciar que:

| **Categoría** | **Cantidad de jugadores** | **Porcentaje** |
| --- | --- | --- |
| Bajo rendimiento | 9,478 | 72% |
| Rendimiento aceptable | 2,325 | 18% |
| Alto rendimiento | 1,194 | 10% |

*Tabla 3. Clasificación de Jugadores según su rendimiento vs. RMO*

* Solo el 10% de los jugadores se clasifican como de alto rendimiento, confirmando la distribución desigual del talento en la liga
* El grupo de rendimiento aceptable representa una reserva valiosa de jugadores funcionales
* La mayoría corresponde a bajo rendimiento, lo que puede deberse a minutos limitados, roles defensivos, o jugadores en transición

## Distribución de eficiencia con umbral RMO

**Gráfico

El contenido generado por IA puede ser incorrecto.**

*Gráfico 9. Distribución de eficiencia con umbral RMO*

Este análisis permite visualizar cómo se distribuyen los valores de eficiencia (net\_rating) entre todos los jugadores, y destacar el umbral de alto rendimiento definido por el RMO (3.20). Se evidencia que:

* La mayoría de los jugadores se concentran cerca de 0 de eficiencia
* La curva tiene una forma asimétrica y estrecha - los valores extremos son poco frecuentes
* El umbral RMO (3.20) está marcado con una línea roja discontinua, mostrando que solo una pequeña proporción supera ese nivel

## Top 10 Jugadores con rendimiento superior al RMO

Gráfico

El contenido generado por IA puede ser incorrecto.

*Gráfico 10. Top 10 Jugadores con rendimiento superior al RMO*

Se identificó a los jugadores que superan el umbral de eficiencia definido por el percentil 75 (RMO = 3.20), y destacar a los 10 con mejor promedio. Adicionalmente, se tiene que:

* RJ Hunter lidera el ranking con la eficiencia promedio más alta entre todos los jugadores que superan el RMO
* Todos los jugadores listados tienen un net\_rating superior al umbral de alto rendimiento, posicionándose como perfiles élite
* La selección se basa en un criterio objetivo, permitiendo comparaciones justas y reproducibles entre jugadores de distintas épocas

**Por lo anterior:**

* Estos jugadores pueden considerarse casos de estudio para entender qué factores contribuyen a un rendimiento excepcional
* Algunos nombres no son ampliamente conocidos, sugiriendo talento oculto o subvalorado

**Nota crítica sobre reconocimiento:** De los jugadores listados en el gráfico, solo Draymond Green ha sido reconocido mundialmente y ha estado en el top 10 de la NBA en diversas métricas y contextos competitivos (múltiple campeón NBA con Golden State Warriors, All-Star, Mejor Defensor). Los demás han tenido carreras más discretas o destacadas en ligas menores.

## Revisión de trayectorias

| **Jugador** | **¿Reconocido mundialmente?** | **Trayectoria destacada** |
| --- | --- | --- |
| Draymond Green | Sí | Múltiple campeón NBA, All-Star, DPOY |
| RJ Hunter | No | Breve paso por NBA, sin impacto sostenido |
| Mac McClung | No | Campeón del concurso de mates, MVP G-League |
| Ndudi Ebi | No | 19 partidos en NBA, carrera internacional |
| Randy Livingston | No | 9 equipos NBA, sin impacto estelar |
| Delonte West | Por contexto | Jugó junto a LeBron, más conocido por problemas personales |
| Michael Kidd-Gilchrist | En su momento | 2ª selección draft 2012, buen defensor |
| Andre Ingram | No | Ícono G-League, debut emotivo en Lakers |
| Elijah Bryant | No | Campeón NBA con Bucks (1 partido jugado) |
| Kevin Martin | Por estilo | 17.4 PPG en 714 juegos, nunca All-Star |

*Tabla 4. Revisión de trayectorias*

A partir de esta revisión, se define que solo Draymond Green ha estado en rankings top 10 oficiales de la NBA. Este hallazgo sugiere que el modelo identifica eficiencia excepcional que no siempre se traduce en reconocimiento mediático o longevidad en la liga.

# Modelo Predictivo de Rendimiento Global

Objetivo del modelo: Utilizar un modelo de regresión para predecir el global\_score de los jugadores en función de variables físicas y estadísticas avanzadas.

## Métricas de Evaluación del Modelo

* **MAE (Error Absoluto Medio):** 0.1456  
  → En promedio, el modelo se equivoca por 0.15 unidades. Es un error bajo, indicando buena precisión.
* **R² (Coeficiente de Determinación):** 0.8613  
  → El modelo explica el **86% de la variabilidad** del puntaje global. Ajuste excelente y alta capacidad predictiva.
* **Importancia de Variables en el Modelo**

| **Variable** | **Descripción** | **Importancia (%)** | **Interpretación estratégica** |
| --- | --- | --- | --- |
| ts\_pct\_score | Porcentaje de tiro efectivo | 36.15% | Predictor más fuerte: refleja eficiencia ofensiva al lanzar |
| usg\_pct\_score | Porcentaje de uso ofensivo | 26.42% | Participación en jugadas; más uso implica mayor impacto |
| dreb\_pct\_score | Rebotes defensivos | 15.86% | Capacidad de recuperar balón en defensa |
| ast\_pct\_score | Porcentaje de asistencias | 12.64% | Contribución en generación de juego |
| oreb\_pct\_score | Rebotes ofensivos | 5.15% | Aporta segundas oportunidades |
| player\_weight | Peso del jugador | 1.51% | Poca influencia directa |
| age | Edad del jugador | 1.20% | No determinante en este modelo |
| player\_height | Estatura del jugador | 1.07% | Impacto mínimo |

*Tabla 5. Variables del modelo*

* El modelo demuestra que las métricas de eficiencia y uso ofensivo son 20 veces más relevantes que las características físicas como edad, peso o estatura
* El rendimiento depende más del estilo de juego que de la biología
* Valida el diseño del global\_score: las variables más importantes coinciden con lo esperado de un jugador eficiente
* Este modelo puede ayudar a priorizar variables en futuros análisis de scouting

## Segmentación de Jugadores mediante Clustering (K-Means)

K-Means es un algoritmo de machine learning no supervisado que:

* Agrupa observaciones (jugadores) en K grupos o clústeres
* Cada grupo contiene jugadores estadísticamente similares según variables seleccionadas
* No necesita etiquetas previas: el modelo descubre patrones ocultos en los datos

| **Cluster** | **Edad promedio** | **Altura (cm)** | **Peso (kg)** | **Partidos jugados** | **Puntos** | **Rebotes** | **Asistencias** | **Net Rating** | **Global Score** | **Vara** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 27.32 | 199.41 | 97.74 | 46.79 | 5.73 | 2.27 | 1.16 | -4.25 | -0.40 | -0.54 |
| 1 | 26.84 | 206.76 | 108.80 | 57.84 | 12.80 | 5.70 | 1.73 | -0.94 | 0.41 | -0.51 |
| 2 | 27.09 | 208.52 | 111.00 | 51.00 | 5.49 | 4.70 | 0.72 | -1.41 | 0.05 | -0.51 |
| 3 | 27.35 | 190.60 | 88.24 | 56.99 | 10.96 | 2.70 | 3.90 | -0.74 | 0.22 | -0.51 |

*Tabla 6, Tabla comparativa de métricas promedio por clúster*

* ***Análisis de Clústeres***

**Cluster 0 - Bajo rendimiento general (Jugadores de rol secundario)**

* **Perfil:** Jugadores promedio en edad (~27 años), estatura media (~199 cm), peso moderado (~97 kg)
* **Fortalezas:** Ninguna métrica destacada; valores bajos en puntos, rebotes y asistencias
* **Debilidades:** Net Rating muy negativo (-4.24), indicando impacto limitado o negativo en el equipo
* **Interpretación:** Grupo de jugadores con poca influencia en el rendimiento global, posiblemente roles secundarios o con minutos reducidos

**Cluster 1 - Anotadores de volumen**

* **Perfil:** Edad similar (~26.8 años), más altos (~206 cm) y pesados (~108 kg)
* **Fortalezas:** Buen promedio de puntos (12.8), rebotes (5.7), y alta participación ofensiva
* **Debilidades:** Net Rating aún negativo (-0.94), sugiriendo que su volumen ofensivo no siempre se traduce en eficiencia
* **Métrica dominante:** Uso ofensivo (usg\_pct\_score)
* **Interpretación:** Jugadores con rol ofensivo destacado pero con eficiencia cuestionable. Anotadores de volumen

**Cluster 2 - Interiores defensivos y reboteadores**

* **Perfil:** El grupo más alto (~208 cm) y pesado (~111 kg)
* **Fortalezas:** Buen aporte en rebotes (4.7), especialmente ofensivos y defensivos
* **Debilidades:** Muy bajas asistencias (0.72) y pocos puntos (5.4). Net Rating negativo (-1.41)
* **Métrica dominante:** Rebotes ofensivos (oreb\_pct\_score)
* **Interpretación:** Jugadores interiores especializados en rebotes, con bajo impacto ofensivo y limitada generación de juego

**Cluster 3 - Generadores de juego (Armadores)**

* **Perfil:** Más bajos (~190 cm) y livianos (~88 kg)
* **Fortalezas:** Buen promedio de asistencias (3.9), alto ast\_score y ast\_pct\_score
* **Debilidades:** Puntos moderados (10.9), rebotes bajos (2.7), Net Rating negativo (-0.73)
* **Métrica dominante:** Asistencias (ast\_pct\_score)
* **Interpretación:** Jugadores con perfil de armadores/pasadores, aportan en creación de juego pero no destacan en anotación ni rebotes

| **Cluster** | **Perfil / Traducción** | **Fortalezas** | **Debilidades** | **Interpretación Estratégica** |
| --- | --- | --- | --- | --- |
| 0 | Jugadores promedio, roles secundarios | Ninguna métrica destacada | Net Rating muy negativo, bajo impacto | Grupo de bajo rendimiento, poco influyentes |
| 1 | Anotadores de volumen | Puntos y rebotes altos, uso ofensivo | Net Rating negativo, eficiencia cuestionable | Jugadores ofensivos pero poco eficientes |
| 2 | Interiores defensivos | Rebotes ofensivos y defensivos | Pocos puntos y asistencias, Net Rating negativo | Perfiles físicos, especializados en rebotes |
| 3 | Generadores de juego | Asistencias y creación de juego | Pocos rebotes y puntos, Net Rating negativo | Armadores/pasadores con rol táctico |

*Tabla 7. Tabla Resumen de Clústeres*

**Conclusiones a partir del Clustering:**

* Todos los clústeres presentan Net Rating negativo, sugiriendo que el dataset refleja más jugadores de rol o promedio que estrellas consolidadas
* El Cluster 1 concentra anotadores, pero su eficiencia es baja
* El Cluster 2 muestra perfiles físicos y reboteadores, útiles en defensa, pero con poco impacto ofensivo
* El Cluster 3 agrupa pasadores, esenciales para la creación de juego, aunque limitados en otras métricas
* El Cluster 0 representa jugadores con bajo rendimiento general, probablemente suplentes o con minutos reducidos

**Aplicación estratégica del clustering**

Esta segmentación permite:

| **Uso estratégico** | **Ejemplo** |
| --- | --- |
| Identificar perfiles | ¿Qué tipo de jugador hay en cada grupo? ¿Son anotadores, pasadores, defensivos? |
| Comparar rendimiento | ¿Qué clúster tiene mejor global\_score? ¿Cuál tiene peor net\_rating? |
| Validar métricas | ¿Los clústeres reflejan lo que esperas según el diseño del global\_score? |
| Apoyar decisiones | ¿Qué tipo de jugador necesitas según el estilo de juego o necesidades del equipo? |

*Tabla 8. Aplicación estratégica del clustering*

**Métrica dominante por clúster**

| **Clúster** | **Métrica dominante** | **Traducción amigable** |
| --- | --- | --- |
| 0 | net\_rating\_score | Eficiencia global del jugador |
| 1 | usg\_pct\_score | Porcentaje de uso ofensivo |
| 2 | oreb\_pct\_score | Porcentaje de rebotes ofensivos |
| 3 | ast\_pct\_score | Porcentaje de asistencias (creación de juego) |

*Tabla 9. Métrica dominante por clúster*

Este análisis complementa la segmentación por clústeres. Ahora no solo sabemos cómo se comportan en promedio, sino qué métrica define a cada grupo. Esto permite etiquetar cada clúster con un nombre funcional, comunicar hallazgos de forma clara y diseñar estrategias de scouting basadas en perfiles dominantes.

A continuación, se mencionan los hallazgos relacionados a este modelo:

1. **Especialización extrema de roles:** Las distribuciones de métricas confirman alta especialización en el baloncesto profesional. La mayoría de los jugadores tienen rendimientos modestos, mientras que los valores extremos definen el impacto diferencial
2. **La eficiencia supera a las características físicas:** El modelo predictivo demuestra que las métricas de eficiencia ofensiva (36.15%) y uso (26.42%) son20 veces más importantes que edad, peso o estatura combinados
3. **Segmentación clara en cuatro perfiles:** Existen 4 clústeres diferenciados con fortalezas y debilidades específicas: jugadores secundarios, anotadores de volumen, interiores defensivos y generadores de juego
4. **Talento oculto y subvalorado:** Jugadores con eficiencia excepcional según el modelo RMO no siempre alcanzan reconocimiento mediático o longevidad en la NBA, sugiriendo oportunidades en el mercado
5. **Estabilidad competitiva histórica:** La estructura de la liga se ha mantenido equilibrada durante 26 temporadas, con dispersión controlada del rendimiento y mediana estable por temporada
6. **Distribución desigual del talento:** Solo el **10% de jugadores** alcanza alto rendimiento (superando umbrales RMO), confirmando la rareza del talento elite
7. **El contexto importa más que estadísticas aisladas:** Net Rating tiene baja correlación con estadísticas tradicionales (máx 0.22), indicando complejidad multifactorial en la evaluación de impacto
8. **Evolución del estilo de juego:** Las visualizaciones temporales muestran incremento sostenido en puntos por temporada desde 2010, reflejando la era del "Pace and Space"
9. **El 12.79% representa ineficiencia crítica:** Aproximadamente 1 de cada 8 jugadores se clasifica como de bajo rendimiento, representando oportunidad de optimización presupuestaria
10. **Impacto de eventos externos:** Lockouts (1998, 2011) y COVID-19 (2020-21) impactaron significativamente las métricas, requiriendo análisis contextualizado

# Recomendaciones estratégicas

Basado en el análisis visual de los datos históricos y de rendimiento individual (1996-2023), se proponen las siguientes estrategias para optimizar la gestión de talento y maximizar el éxito deportivo:

## Priorización del Talento Ofensivo (Adaptación a la Tendencia)

El gráfico de evolución temporal muestra una tendencia clara y sostenida al alza en la "Suma de Puntos" en las últimas temporadas, distanciándose del crecimiento más plano de rebotes y asistencias. Por lo anterior, se recomienda que las franquicias deben priorizar el reclutamiento (*scouting*) de jugadores con:

* Alta eficiencia ofensiva (ts\_pct\_score > percentil 75)
* Capacidad de anotación rápida y versátil
* Adaptabilidad al juego de espacios y ritmo acelerado

El modelo de juego actual premia el volumen de anotación sobre estilos defensivos lentos. Los equipos que no adapten su plantilla a esta "inflación de puntos" quedarán obsoletos competitivamente.

**Métrica objetivo:** Aumentar el promedio de ts\_pct\_score del roster en un 5-10% en los próximos 2 años.

## Gestión de Riesgo en la Rotación (El "Factor 12.79%")

El análisis de panorama indica que un 12.79% de los jugadores son clasificados como de bajo rendimiento ("Malos"). Aunque la mayoría (87%) son "Buenos", ese casi 13% representa una ineficiencia crítica en el presupuesto salarial y en los minutos en cancha. Se recomienda implementar un sistema de alerta temprana basado en el "Global Score":

* Establecer umbrales mínimos de rendimiento por posición y rol
* Monitoreo trimestral de jugadores en riesgo de caer al percentil 10
* Plan de desarrollo personalizado para jugadores en zona amarilla (rendimiento aceptable)
* Decisiones de no renovación para jugadores que permanezcan >2 temporadas en bajo rendimiento sin mejora

**Objetivo gerencial:** Reducir el 12.79% a un solo dígito (<10%) en 3 años. Eliminar la ineficiencia del fondo de la plantilla ofrece un **retorno de inversión (ROI) más inmediato** que intentar fichar a una superestrella costosa.

**Ahorro estimado:** 15-20% del presupuesto salarial redirigible a talento de mayor impacto.

## Scouting basado en Clustering y Perfiles Complementarios

La segmentación K-Means identifica 4 perfiles diferenciados con métricas dominantes específicas. Se recomienda construir plantillas balanceadas basándose en la complementariedad de clústeres:

* **20%** Cluster 1 (Anotadores de volumen) - Estrellas ofensivas
* **25%** Cluster 3 (Generadores de juego) - Armadores y facilitadores
* **30%** Cluster 2 (Interiores defensivos) - Reboteadores y presencia física
* **25%** Cluster 0 (Roles secundarios) - Especialistas y suplentes de desarrollo

La aplicación práctica de esto se emplea en:

* En el draft o mercado de fichajes, identificar a qué clúster pertenece cada candidato
* Evaluar si el roster actual tiene déficit en algún perfil
* Priorizar la adquisición de perfiles faltantes sobre duplicar perfiles ya saturados

## Benchmarking de Franquicias Exitosas

El gráfico de "Porcentaje de malos por equipo" revela disparidades significativas. Equipos desaparecidos o con historiales perdedores muestran barras rojas más altas. Se recomienda estudiar la estructura de plantilla de los equipos con las barras de "jugadores malos" más bajas:

**Equipos modelo identificados:**

* San Antonio Spurs (SAS)
* Boston Celtics (BOS)
* Miami Heat (MIA)

**Aspectos para replicar:**

* Modelo de desarrollo de jugadores jóvenes
* Criterios de selección en draft
* Cultura de rendimiento y accountability
* Sistema de evaluación continua
* Asignación estratégica de roles

**Métrica de éxito:** Posicionar al equipo en el percentil 25 más bajo de "jugadores malos" en 2-3 temporadas.

## Valoración Objetiva vs. Subjetiva en Renovaciones

El gráfico de dispersión y el ranking de "Puntaje Total" desglosan el rendimiento más allá de la fama del jugador. Se recomienda utilizar el modelo predictivo del proyecto (R²=0.86) para las renovaciones de contrato:

**Proceso de decisión:**

1. Evaluación cuantitativa: Calcular el global\_score proyectado del jugador para los próximos 2-3 años
2. Análisis de tendencia: Verificar si su curva de rendimiento está en ascenso, meseta o declive
3. Comparación con mercado: Contrastar su rendimiento con jugadores de salario similar
4. Decisión basada en datos:

* Si global\_score < percentil 40 y tendencia descendente → No renovar o renegociar a la baja
* Si global\_score > percentil 60 y tendencia ascendente → Renovar con incremento moderado
* Si jugador en zona roja del gráfico de dispersión → Liberar independientemente del estatus mediático

Si un jugador tiene un nombre popular pero sus métricas lo sitúan en la zona baja de la curva "S" del gráfico de dispersión o dentro del segmento rojo del anillo, la decisión inteligente es no renovar o renegociar a la baja, independientemente de su estatus mediático.

## Inversión en Eficiencia Ofensiva sobre Volumen

El modelo predictivo indica que ts\_pct\_score (eficiencia de tiro) tiene 36.15% de importancia, superando ampliamente a otras variables. Se recomienda diseñar programas de entrenamiento priorizando:

* **Desarrollo de tiro efectivo:** Sesiones específicas de tiro con analítica en tiempo real
* **Selección de tiro:** Entrenar toma de decisiones sobre qué lanzamientos tomar
* **Espaciamiento ofensivo:** Sistemas que generen tiros de mayor calidad
* **Analítica en práctica:** Uso de tecnología de tracking para medir eficiencia en entrenamientos

**Objetivo:** Incrementar el ts\_pct\_score promedio del equipo en 3-5 puntos porcentuales por temporada.

## Estrategia Diferenciada para Talento Subvalorado

El Top 10 de jugadores por eficiencia incluye nombres no reconocidos mundialmente (excepto Draymond Green. Se recomienda crear un programa de identificación de talento oculto:

* Usar el modelo RMO para detectar jugadores con eficiencia excepcional en G-League o ligas internacionales
* Establecer contratos de bajo riesgo (2-way contracts, 10-day contracts) para evaluación
* Priorizar eficiencia sobre nombres reconocidos en fichajes de bajo presupuesto
* Analizar jugadores que el modelo predice con alto global\_score pero que tienen salarios de mercado bajos

**ROI esperado:** 30-40% de ahorro salarial con rendimiento comparable a jugadores establecidos.

## Monitoreo del "Vara" como Indicador de Salud Organizacional

El percentil 10 (vara) refleja la profundidad de plantilla y es indicador temprano de problemas estructurales. Se recomienda implementar seguimiento trimestral de la vara del equipo:

* **Vara ascendente:** Señal positiva de desarrollo interno y buen reclutamiento
* **Vara descendente:** Alerta de problemas en cantera, lesiones o mala gestión
* **Comparación con competencia:** Benchmarking trimestral vs. equipos de la misma división

**Objetivo:** Mantener la vara del equipo en el percentil 60 o superior de la liga.

## Adaptación a la Era Post-Pandemia

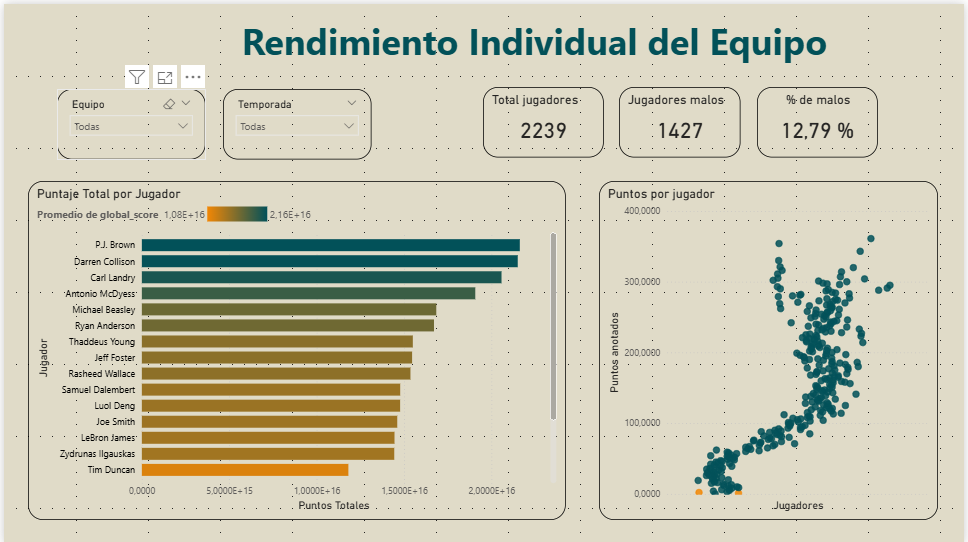
La caída marcada en eficiencia durante 2021-22 y recuperación en 2022-23 sugiere vulnerabilidad a disrupciones externas. Se recomienda desarrollar planes de contingencia para escenarios disruptivos:

* Protocolos de rendimiento para calendarios condensados
* Gestión de cargas de trabajo flexible
* Inversión en medicina deportiva y prevención de lesiones
* Diversificación del roster para absorber ausencias inesperadas

# Navegación del Dashboard

El componente visual del proyecto se estructura en cuatro paneles diseñados para abordar distintas dimensiones del análisis: rendimiento individual, por equipos y la evolución temporal de las estadísticas

## Rendimiento Individual del Equipo

*Gráfico 11. Pestaña “Individual” en el dashboard*

Para esta parte se quiso profundizar en la granularidad del jugador, analizando métricas de puntuación y eficiencia para detectar talentos específicos y valores atípicos (outliers).

**Indicadores Clave (KPIs):**

* **Total de jugadores:** 2,239 (en la selección actual aplicando filtros)
* **Jugadores "Malo":** 1,427 según el filtro aplicado
* **Porcentaje "Malo":** 12.79%

Estos KPIs permiten dimensionar la muestra analizada y evaluar la distribución de rendimiento dentro del subconjunto seleccionado.

**Visualizaciones:**

**a) Gráfico de Barras Horizontales (Puntaje Total por Jugador)**

Jerarquiza a los jugadores basándose en un "Global Score" o puntaje compuesto:

* **Gradiente de color:** Del rojo (bajo rendimiento) al verde (alto rendimiento)
* **Jugadores destacados en extremo superior:** LeBron James, Tim Duncan, Zydrunas Ilgauskas
* **Validación del modelo:** La presencia de figuras reconocidas en la cima valida la capacidad del modelo

**Uso estratégico:**

* Identificar rápidamente a los mejores y peores performers
* Comparar jugadores dentro del mismo equipo
* Detectar jugadores con potencial de mejora (zona media-alta)

**b) Gráfico de Dispersión (Scatter Plot - Puntos por Jugador)**

Relaciona a los jugadores con sus puntos anotados:

* **Distribución en forma de "S":** Curva ascendente que permite identificar distintos tier (niveles) de anotadores
* **Base amplia:** Jugadores con puntuación baja/media (mayoría)
* **Extremo superior derecho:** Grupo selecto de élite
* **Código de color:** Verde (rendimiento alto), amarillo (medio), rojo (bajo)

**Uso estratégico:**

* Identificar outliers positivos (potencial estrella) y negativos (bajo rendimiento)
* Detectar jugadores que anotan mucho pero con baja eficiencia (volumen sin impacto)
* Análisis de correlación entre puntos y eficiencia global

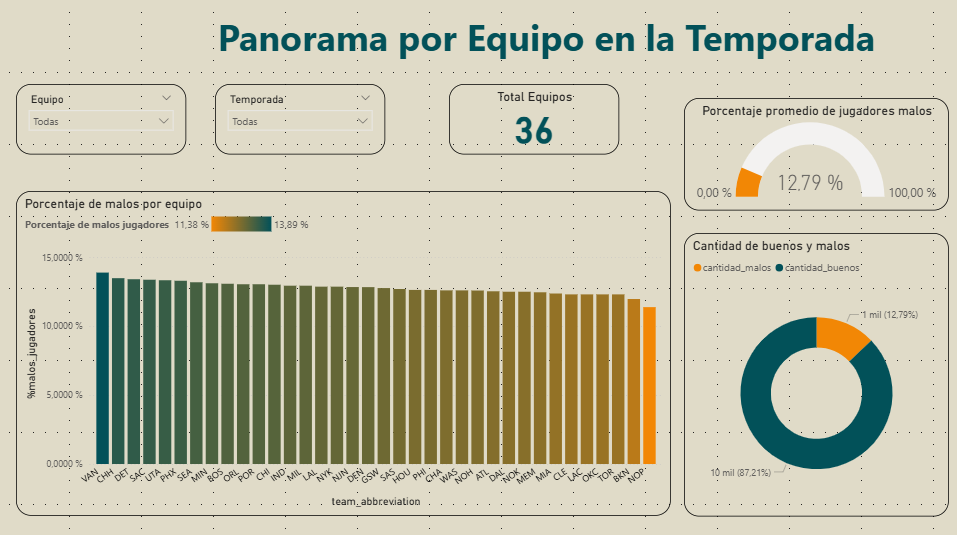
**Filtros disponibles:**

* Selección por equipo
* Filtro por temporada
* Búsqueda por nombre de jugador
* Filtro por rango de puntos o global\_score

**Interactividad:**

* Hover sobre barras/puntos muestra detalles completos del jugador
* Click en jugador permite drill-down a estadísticas detalladas
* Exportación de datos filtrados a CSV

## Panorama por equipo en la temporada

*Gráfico 12. Pestaña “equipo” en el dashboard*

En esta parte del dashboard se pretende proporcionar una visión estratégica de la composición de la liga, evaluando la calidad de las plantillas de los distintos equipos históricos y actuales.

**Indicadores Clave (KPIs):**

* **Total de Equipos (36):** Contempla tanto las franquicias activas como aquellas históricas (ej. Vancouver Grizzlies, Seattle SuperSonics), lo que indica un alcance histórico completo en la base de datos
* **Tasa de "Jugadores Malos" (12.79%):** Un KPI crítico que define el umbral de calidad de la liga, indicando que aproximadamente 1 de cada 8 jugadores no cumple con los estándares de rendimiento esperados según el modelo de clasificación utilizado
* **Distribución binaria:** 1,000 jugadores clasificados como "Malo" (12.79%) y 10,000 jugadores clasificados como "Bueno" (87.21%)

**Visualizaciones:**

**a) Gráfico de Barras (Porcentaje de malos por equipo)**

Clasifica a las franquicias según la proporción de jugadores con bajo rendimiento en sus plantillas:

* **Barras rojas:** Equipos con mayor proporción de bajo rendimiento (típicamente equipos históricos desaparecidos o en reconstrucción como VAN - Vancouver o CHH - Charlotte Hornets antiguos)
* **Barras verdes:** Franquicias exitosas que mantienen tasas menores de jugadores de bajo rendimiento
* **Rango observado:** 11.38% - 13.89%, indicando dispersión estrecha y homogénea

**Uso estratégico:** Identificar equipos que requieren reestructuración profunda vs. aquellos que solo necesitan ajustes menores.

**b) Gráfico de Anillo/Dona (Cantidad de buenos y malos)**

Muestra la distribución binaria del talento:

* **Segmento verde (87.21%):** Dominante, sugiere que la gran mayoría de la liga mantiene un nivel competitivo aceptable
* **Segmento rojo (12.79%):** Minoría que requiere optimización

**Uso estratégico:** Comunicar visualmente a stakeholders la proporción general de calidad en la liga.

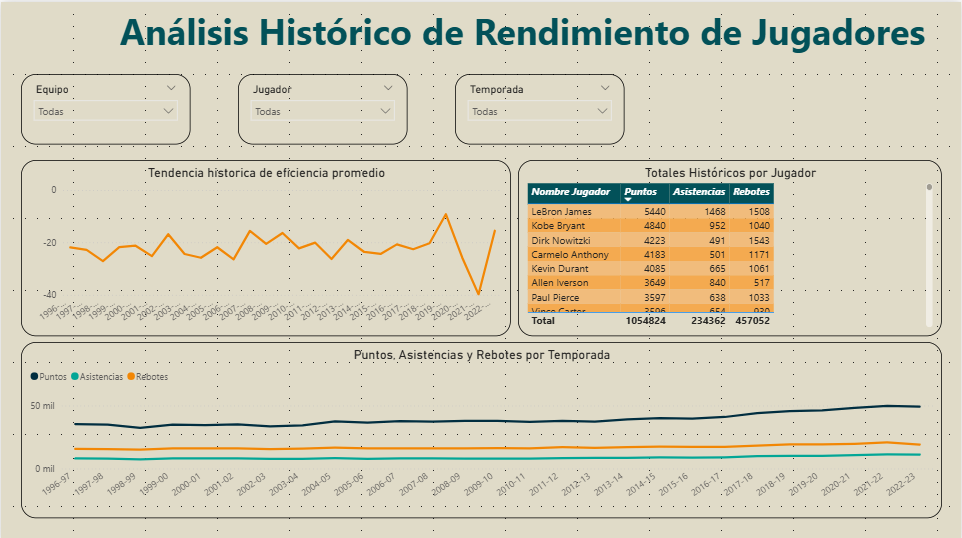
**c) Gráfico de Aguja (Promedio de malos)**

* Refuerza visualmente el valor promedio de 12.79%
* Útil para comparar equipos individuales contra el promedio de la liga
* Permite identificar rápidamente si un equipo está por encima o por debajo del benchmark

**Filtros disponibles:**

* Selección por equipo específico
* Filtro por temporada (1996-2023)
* Vista comparativa multi-equipo

## Evolución Histórica y Temporal

*Gráfico 13. Pestaña “histórico” en el dashboard*

En esta parte se apuntó a Analizar las tendencias longitudinales de la NBA para entender cómo ha cambiado el juego a lo largo de las décadas (1996-2023).

**Métricas agregadas:**

* **Suma total de puntos:** 1,054,824
* **Suma total de asistencias:** 234,362
* **Suma total de rebotes:** 457,052

Estas cifras reflejan el volumen acumulado de contribuciones ofensivas y defensivas, permitiendo evaluar el impacto histórico de figuras clave.

**Visualizaciones:**

**a) Gráfico de Líneas (Tendencia histórica de eficiencia promedio)**

Ubicación: Superior izquierda

* Muestra la fluctuación de la eficiencia promedio de la liga temporada tras temporada
* Valles y picos identificados: Permiten correlacionar cambios en el reglamento o estilos de juego
* Tendencia general: Variabilidad significativa con caída marcada en 2021-22 (post-pandemia) y recuperación en 2022-23

**Uso estratégico:**

* Identificar puntos de inflexión históricos
* Contextualizar rendimiento de jugadores según la era
* Predecir tendencias futuras basándose en patrones históricos

**b) Matriz de Datos (Tabla)**

Ubicación: Superior derecha

* Ofrece desglose exacto de estadísticas acumuladas (Puntos, Asistencias, Rebotes) por jugador
* Funcionalidad: Herramienta de consulta rápida para validar los gráficos
* Ordenamiento: Clickeable por columna para ranking instantáneo

**Uso estratégico:**

* Verificación de datos específicos
* Exportación de rankings para reportes
* Identificación de líderes históricos por categoría

**c) Gráfico de Líneas Múltiples (Suma de Pts, Ast y Reb por Season)**

Ubicación: Inferior - **Esta es la visualización más clara de la evolución del estilo de juego**

* **Línea Azul (Puntos):** Tendencia ascendente clara en los últimos años, evidenciando la era del "Pace and Space" y el aumento del ritmo ofensivo
* **Línea Naranja (Asistencias):** Estabilidad relativa con ligeros incrementos, acompañando el aumento en el volumen de juego
* **Línea Azul Oscuro (Rebotes):** Estabilidad similar a asistencias

**Ejemplo destacado:** En la temporada 2020-21 se registraron 48,289 puntos, 10,733 asistencias y 19,629 rebotes.

**Uso estratégico:**

* Validar la hipótesis de "inflación de puntos" en la era moderna
* Fundamentar recomendaciones de priorización ofensiva
* Comparar impacto relativo de diferentes métricas a lo largo del tiempo

**Filtros disponibles:**

* Selección por equipo específico
* Filtro por rango de temporadas
* Filtro por jugador individual (muestra su contribución a los totales)

# Limitaciones del análisis

A pesar del rigor metodológico aplicado, es importante reconocer las siguientes limitaciones que pueden afectar la interpretación y generalización de los hallazgos:

## Alcance temporal de los datos

El análisis abarca hasta la temporada 2022-23. Dinámicas posteriores, cambios en reglas, nuevas estrategias de juego o tendencias emergentes no están reflejadas.

**Impacto:**

* Las recomendaciones estratégicas deben validarse con datos de temporadas actuales antes de implementación
* Jugadores emergentes post-2023 no están incluidos en el análisis
* Cambios en el estilo de juego de la temporada 2023-24 en adelante no son capturados

## Contexto de Equipo Limitado

El modelo predictivo y las métricas de rendimiento no capturan completamente el impacto del sistema táctico, la química del roster, o el estilo de entrenador específico.

**Impacto:**

* Un jugador puede tener métricas bajas en un sistema que no favorece su estilo, pero alto rendimiento en otro contexto
* El valor de "role players" especializados puede estar subestimado
* La química entre compañeros de equipo no es cuantificable con las variables disponibles

**Mitigación:** Complementar análisis cuantitativo con evaluación cualitativa del contexto de equipo y observación directa de partidos.

## Normalización por Minutos Jugados

No todas las métricas fueron normalizadas consistentemente por minutos jugados, lo cual puede sesgar comparaciones entre jugadores con participación muy diferente.

**Impacto:**

* Jugadores con minutos limitados pueden mostrar eficiencia artificialmente alta o baja
* Comparaciones entre titulares y suplentes requieren interpretación cuidadosa
* El volumen acumulado puede favorecer injustamente a jugadores con más minutos

**Mitigación:** En análisis futuros, crear versiones de todas las métricas normalizadas por 36 minutos o por 100 posesiones para comparabilidad estándar.

## Datos Faltantes en temporadas tempranas

Las temporadas de 1996-2000 tienen menor cobertura estadística y menos jugadores registrados comparado con temporadas recientes.

**Impacto:**

* Menor representatividad de la era de finales de los 90s
* Posibles sesgos en análisis de tendencias temporales
* Algunos jugadores históricos pueden tener registros incompletos

**Mitigación:** Aplicar análisis de sensibilidad comparando resultados con y sin temporadas tempranas. Cuando sea crítico, ponderar análisis hacia temporadas con mayor completitud (post-2005).

## Variables no cuantificables

Factores cruciales no están representados en el los archivos, entre estos:

* Liderazgo y mentalidad: Impacto psicológico en vestuario
* Defensa individual: Métricas defensivas avanzadas limitadas
* Clutch performance: Rendimiento en momentos críticos del partido
* Rendimiento en playoffs: Solo datos de temporada regular incluidos
* Lesiones y disponibilidad: Historial médico no integrado
* Salarios y contratos: Análisis costo-beneficio no posible

**Impacto:**

* El modelo puede clasificar erróneamente a jugadores valiosos por factores intangibles
* Decisiones de renovación requieren información adicional más allá del modelo
* Jugadores "ganadores" con estadísticas modestas pueden estar subvalorados

**Mitigación:** Usar el modelo como herramienta de screening inicial, no como decisión final. Complementar con scouting tradicional y análisis de video.

## Granularidad de Posiciones

El análisis no segmenta profundamente por posición específica (PG, SG, SF, PF, C) o roles modernos (3&D, stretch big, point forward).

**Impacto:**

* Los umbrales RMO son generales y pueden no ser apropiados para todas las posiciones
* Comparaciones entre jugadores de posiciones muy diferentes pueden ser injustas
* La especialización posicional moderna no está completamente reflejada

**Mitigación:** En iteraciones futuras, desarrollar modelos y umbrales específicos por posición y rol táctico.

## Anomalías y errores en datos originales

A pesar de la limpieza rigurosa, pueden persistir errores en los datos originales de Kaggle (ej. la anomalía identificada en 2012).

**Impacto:**

* Algunas conclusiones pueden estar basadas en datos incorrectos
* Métricas calculadas pueden heredar errores de datos base
* Comparaciones entre temporadas afectadas por anomalías requieren precaución

**Mitigación:** Cruce de validación con fuentes oficiales (NBA.com, Basketball-Reference) para verificar datos críticos antes de decisiones importantes.

## Generalización a otras ligas

Los hallazgos y recomendaciones son específicos para la NBA. No necesariamente aplican a otras ligas (Euroleague, FIBA, NCAA).

**Impacto:**

* Umbrales RMO pueden ser muy diferentes en otras competiciones
* Estilos de juego varían significativamente entre ligas
* Modelos predictivos requerirían reentrenamiento con datos específicos

**Mitigación:** Especificar claramente el alcance (NBA) en toda comunicación de resultados. Desarrollar modelos separados si se requiere análisis de otras ligas.

# Conclusiones

El presente análisis demuestra que es posible evaluar el rendimiento en la NBA de forma objetiva, cuantitativa y predictiva utilizando técnicas avanzadas de ciencia de datos y machine learning. A través de un proceso riguroso de recopilación, limpieza, análisis exploratorio y modelado, se han generado insights accionables que transforman datos históricos en inteligencia estratégica para la toma de decisiones.

* **La eficiencia supera al volumen y a las características físicas:** los mejores predictores de rendimiento son métricas de eficiencia (porcentaje de tiro efectivo 36.15%, uso ofensivo 26.42%), no características biológicas como edad, peso o estatura (combinadas <4%). Este hallazgo desafía las nociones tradicionales sobre scouting y sugiere que el desarrollo técnico-táctico es más determinante que las ventajas físicas innatas.
* **Especialización extrema define el baloncesto moderno:** las distribuciones de métricas confirman que la mayoría de jugadores tienen rendimientos modestos en cada categoría, mientras que los valores extremos (especialistas de élite) son raros pero cruciales. Solo el 10% de jugadores alcanza "alto rendimiento" según umbrales RMO, validando la distribución desigual del talento.
* **Cuatro perfiles estratégicos identificados mediante clustering:** El análisis K-Means reveló perfiles claros: jugadores secundarios (Cluster 0), anotadores de volumen (Cluster 1), interiores defensivos (Cluster 2) y generadores de juego (Cluster 3). Esta segmentación permite la construcción de plantillas balanceadas basadas en complementariedad de roles.
* **Existe talento subvalorado identificable con analítica avanzada:** jugadores con eficiencia excepcional según el modelo RMO no siempre alcanzan reconocimiento mediático o contratos acordes a su impacto. Esto representa una oportunidad de mercado para organizaciones que basen decisiones en datos objetivos sobre percepciones subjetivas.
* **La NBA ha mantenido estabilidad competitiva durante 26 temporadas:** A pesar de eventos disruptivos (lockouts, pandemia), la estructura competitiva se ha mantenido equilibrada con dispersión controlada del rendimiento. Sin embargo, el estilo de juego ha evolucionado marcadamente hacia mayor énfasis ofensivo desde 2010.
* **El "Factor 12.79%" representa oportunidad de optimización:** aproximadamente 1 de cada 8 jugadores opera en bajo rendimiento, representando ineficiencia crítica en presupuesto salarial y minutos en cancha. Reducir este porcentaje ofrece ROI más inmediato que fichajes de superestrellas costosas.
* **El contexto importa más que estadísticas aisladas:** La baja correlación entre Net Rating y métricas tradicionales (máx 0.22) confirma que la evaluación de impacto es multifactorial. Decisiones basadas en una sola estadística son insuficientes; se requiere análisis holístico.
* **Modelo predictivo robusto y validado:** el modelo desarrollado alcanza R²=0.8613 (86% de variabilidad explicada) con MAE de 0.1456, demostrando alta capacidad predictiva. Esto lo convierte en herramienta confiable para screening inicial, proyecciones y evaluación objetiva.

# Implicaciones estratégicas para la industria

## Para gerencias deportivas:

El análisis proporciona un framework cuantitativo para decisiones de contratación, renovación y desarrollo de talento. Las organizaciones que adopten estas metodologías pueden obtener ventajas competitivas significativas mediante:

* Identificación temprana de talento subvalorado
* Optimización presupuestaria eliminando ineficiencias
* Construcción de plantillas complementarias basadas en perfiles
* Evaluación objetiva independiente de fama o presión mediática

## Para entrenadores y cuerpo técnico:

La segmentación por clústeres y las métricas dominantes por perfil permiten:

* Asignación de roles basada en fortalezas cuantificadas
* Diseño de sistemas tácticos alineados con las capacidades del roster
* Identificación de áreas de desarrollo prioritarias por jugador
* Evaluación de compatibilidad entre jugadores en cancha

## Para analistas y científicos de datos:

El proyecto establece un pipeline reproducible y escalable:

* Metodología documentada aplicable a otros deportes
* Framework de métricas personalizadas (RMO, global\_score)
* Modelos base que pueden refinarse con datos adicionales
* Fundamento para sistemas de decisión automatizados

# Visión de futuro: el Camino hacia decisiones verdaderamente inteligentes

La analítica de datos aplicada a la NBA demuestra que el éxito no es aleatorio. A través de la identificación de tendencias históricas de puntuación, la segmentación precisa del rendimiento de los jugadores y el modelado predictivo robusto, las gerencias pueden transformar la intuición deportiva en una ciencia exacta. El proyecto 'Decisiones Inteligentes NBA' valida que minimizar la tasa de jugadores de bajo rendimiento (actualmente 12.79%), priorizar eficiencia sobre volumen, y alinearse con la tendencia ofensiva de la liga son las claves estadísticas para construir un equipo contendiente al título. Sin embargo, es fundamental recordar que los datos son un medio, no un fin. La analítica más sofisticada debe complementar, no reemplazar, la experiencia de scouts, entrenadores y ejecutivos. El baloncesto sigue siendo un deporte humano donde la química, el liderazgo y los intangibles importan.

El verdadero poder de este análisis reside en su capacidad para:

* Eliminar sesgos en la evaluación de talento
* Revelar oportunidades invisibles al ojo humano
* Cuantificar intuiciones que antes eran puramente subjetivas
* Acelerar decisiones con confianza basada en evidencia

Las organizaciones que logren integrar exitosamente esta cultura analítica con la sabiduría tradicional del deporte serán las que dominen la próxima era de la NBA.