Análisis de estadísticas de NBA

Máster data analytics semipresencial

Bryan Eduardo De Sousa Velásquez

Juan García Bericua Goyanes

Nuria Heras Louzao

Victor Maestro Calvo

Alberto Sánchez Delgado

2020

INDICE

# INTRODUCCIÓN

# ANTECEDENTES

# HIPOTESIS

# METODOLOGIA DE TRABAJO

# COMO HA EVOLUCIONADO EL BASKET. EVOLUCION GRAFICA

# MODELO DE EQUIPO/JUGADOR PARA SABER CUAL ES LA MEJOR VARIABLE

# MODELO PARA SABER QUIEN VA GANAR

# RESPUESTAS A LAS HIPOTESIS

# REFERENCIAS BIBLIOGRÁFICAS / WEBGRAFÍA

* Renato Amorim Torres. ***Prediction of NBA games based on Machine Learning Methods.***(2013) <http://homepages.cae.wisc.edu/~ece539/fall13/project/AmorimTorres_rpt.pdf>.
* Matthew Beckler, Hongfei Wang. **NBA Oracle** (<http://www.mbeckler.org/coursework/2008-2009/10701_report.pdf> )
* <https://towardsdatascience.com/predicting-2018-19-nbas-most-valuable-player-using-machine-learning-512e577032e3>
* <https://www.basketball-reference.com/>
* SERRANO y BAKER, Basketball Analytics, 2013, Advanced Metrics, Lake St. Louis
* <https://www.basketball-reference.com/leagues/NBA_2014.html#all_team_stats>
* <http://sportsmadeinusa.com/baloncesto/nba/ha-cambiado-el-modo-de-jugar-nba/>
* García, J. Ibáñez, S. J., Feu, S. (2010). Estadísticas de juego que discriminan las selecciones participantes en los Campeonatos del Mundo de 2006 de Baloncesto, en función del nivel y género de los equipos. Revista Kronos, IX, 17, 57-63.
* Sampaio, J., Fraga, F., Silva, T. (2004). La evolución de las estadísticas de los tiros libres en partidos de baloncesto de formación. III Congreso de la Asociación Española de Ciencias del Deporte. Valencia, 11-13 de marzo.
* Sampaio, J., Leite, N. (2006). ¿Por qué ganaron o perdieron los partidos de baloncesto los equipos que participaron en el Eurobasket 2005? Revista Kronos, nº9, pp.67-73.
* Sampaio, J., Lorenzo, A. Gómez Ruano, M. A., Matalarranha, J., Ibáñez, S. I., Ortega, E. (2009). Análisis de las estadísticas discriminantes en jugadores de baloncesto según su puesto específico, en las finales de las competiciones europeas (1988- 2006). Diferencias entre jugadores titulares y suplentes. Apunts, 96, 53-58.

*“La evaluación del jugador de baloncesto en los partidos es un proceso que despierta mucha atención de los entrenadores e investigadores” (Sampaio, 2002).*

# INTRODUCCIÓN

En un mundo en el que hasta ahora hemos basado nuestras decisiones empresariales y personales en experiencias y estadísticas de pequeños conjuntos de información, hemos pasado a un mundo en el que disponemos de gran cantidad de información.

La era actual de tecnologías de la información y comunicación, ha permitido el registro de grandes cantidades de datos para posteriormente manipular la información y sacar conclusiones sobre la conducta de los seres humanos, ventajas para empresas y con aplicaciones para todos los sectores de la sociedad.

El análisis de grandes cantidades de datos para convertirlos en información relevante resulta una tarea compleja, pero que se hay ido aplicando a todos los aspectos de la vida, siendo el deporte uno de los que más ventajas se han obtenido, especialmente de deportes de táctica y estadísticas como es el baloncesto.

La NBA está en constante evolución. El estilo de juego, las capacidades de los jugadores, la información de los equipos, son variables que han ido modificando el estilo de juego, hacia un estilo más defensivo y físico, en el que prima la planificación y la innovación, no tiene cabida los jugadores y entrenadores que son absorbidos por la gran cantidad de información que se extrae en cada partido. Dado este paradigma, nos preguntamos, ¿Cómo ha evolucionado el estilo de juego?( Intro a la parte de Nuria)

Dentro de los datos obtenidos de cada equipo, como influye cada jugador en computo de juego global del equipo. ¿Los grandes jugadores siguen teniendo la influencia en el juego que tenían antes? Cómo saben las franquciias cual es el estilo de jugador que mejor se adapta a las condiciones del equipo?

La búsqueda de una variable que cuantifique esta métrica, es uno de los objetivos que intentaremos obtener con el análisis de los datos en este trabajo ( Intro a la parte de Alberto)

El conocimiento del estilo de juego y de qué jugador es más óptimo para nuestro estilo de juego nos llega a preguntarnos si es posible la proyección de un modelos que nos permite averiguar las probabilidades y derrotas. Esta variable la hemos aplicado junto con las distancias que hay entre los campos de cada equipo y la influencia de los lesionados en esta.

# ANTECEDENTES

Todos los entrenadores e investigadores han buscado la forma de conseguir sacarle el máximo rendimiento a las estadísticas que se recogen durante el tiempo del partido, tanto de sus propios jugadores como del resto de equipos de la NBA. Dentro de los avances obtenidos en el desarrollo y aplicación de nuevos métodos para el análisis cuantitativo de datos deportivos, sobre todo utilizando técnicas estadísticas. Esa acumulación incalculable de datos requiere de conclusiones que nos permitan analizar el estado de nuestro equipo, valorar el impacto de nuestros jugadores dentro del colectivo, prever futuras líneas de actuación y hasta cambiar puntualmente el estilo de juego en función de las estadísticas propias y del equipo rival

Desde la implementación hace más de un lustro de la medición, utilización y comprensión de los datos de estadísticas convencionales como, puntos, rebotes, asistencias nos ha permitido ir conociendo cómo ha evolucionado el juego.

La incorporación de tecnología para recabar datos se ha ido realizado paulatinamente, aunque el mayor avance se produce en el año 2013, con la instalación de cámaras de registro de movimientos, para proporcionar a los entrenadores, jugadores y aficionados la información de todos los partidos.

Los principales estudios se centran en el **ritmo de juego** (se centra en el uso de las faltas, las posesiones, y las jugadas desarrolladas durante el tiempo de partido), los **porcentajes de tiro** (desarrollo de un algoritmo para conocer le valor medio de los tiros en el campo, contando con un valor extra para el lanzamiento de línea de tres) y el intento de conocer mediante el análisis de datos cual es el **mejor ataque y la mejor defensa.**

Siendo estos los temas principales, se han desarrollado varios estudios acerca de las pérdidas de balón y varios estudios sobre las votaciones para los jugadores en el draf y la evolución del juego de triples.

Entre los estudios que destacan por su notoriedad se encuentran, el estudio de Ibáñez y Col y Sampaio y Janeira ambos del 2003 en el que sus conclusiones es que los equipos que anotan más tiros de 2 puntos y recogen más balones debajo del arco obtienen un índice mayor de partidos ganados.

La tesis de Sampaio y Leite, recoge los datos del Eurobasket d 2005, en el que concluyen que las estadísticas más diferenciadora se situaba en la línea d tres, los puntos fallados y las asistencias. Con este estudio se planteaba que la situación ofensiva se concluía con un tiro cercano, con elevadas probabilidades de aciero, para minimizar los riesgos del tiro exterior y escogiendo las mejores opciones de lanzamiento.

un estudio de Hierro (2002), tras analizar la liga ACB y la NBA, concluye que los equipos con mejores porcentajes de tiros de 2 y de 3, mayor número de faltas recibidas, tapones y asistencias, obtenían más victorias; no considerando el número de pérdidas de balón y los tiros libres como aspectos definitivos.

No todos los estudios se centran en las estadisiticas, ya que algunos toman de referencia la edad de los jugadores en la canga, (García, Ibañez y Feu (2007)), así como encontrar la posición más influyente en los partidos. Este estudio (Sampaio et Al, 2009), determinó que los jugadores más importantes eran los bases titulares y los pívots suplentes, reflejándose su aportación en los lanzamientos de 3 puntos, los rebotes y los lanzamientos de 2 puntos.

Otra de las líneas de investigación que hemos encontrado en el análisis de la literatura científica, es el estudio de las victorias y las derrotas en función de la condición del equipo como local o visitante. En la mayoría de casos, los equipos que juegan en casa, tienen mayor ventaja (Courneya y Carron, 1992; Sánchez et al., 2001; Smith, 2005), debido a factores como viajes realizados, el conocimiento de la cancha de juego, la presión percibida por los árbitros, la familiaridad con el estadio, el público y determinadas variables psicológicas

Diversos estudios cuentan con bbdd obtenidas de páginas de datos gratuitas, pero que en nuestro caso, hemos querido sacar nuestras propias variables, para así enriquecer este estudio. El objetico es la creación de un modelo que contenga como base la información más completa de estadísticas de los jugadores a nivel individual y como equipo.

# HIPOTESIS

Con la presentación de este trabajo final de Data analytics nuestro objetivo es la creación de un modelo predictivo que nos ayude a identificar que equipo va a ganar un partido durante el campeonato.

Sabemos que durante el desarrollo de un partido intervienen una gran cantidad de variables que puedan hacer cambiar el resultado final de un partido. Lesiones, acierto de tiros, faltas cometidas son algunas de las variables más importante que pueden llegar a afectar a los partidos.

Dentro de conocer, la importancia para escoger estas métricas hemos desarrollado un modelo que nos explique cual ha sido la evolución del baloncesto en los últimos años, intentado descifrar mediante los datos como ha evolucionado cada una de estas métricas y su influencia en los resultados.

Además

# METODOLOGÍA DE TRABAJO

Para el desarrollo de nuestro planteamiento hemos realizado un scrapping a distintas webs, para obtener la información que necesitábamos para el desarrollo del modelo. El objetivo de este desarrollo ha sido el de enriquecer nuestros dataframes y recopilar información útil, así cono desechar aquellos datos que no eran necesarios.

Para la extracción hemos utilizado Python empleando librerías request, para realizar las peticiones y Pandas para recopilar la información. Las webs que hemos utilizado para sacar los daros han sido [www.basketballreference.com](http://www.basketballreference.com), así como otras webs secundarias como data.nba.net, [www.shrpsports.com](http://www.shrpsports.com) y [www.geodatos.net](http://www.geodatos.net).

Durante el proceso de extracción de datos, se han obtenido varios documentos que pasamos a enumerar, haciendo un breve análisis del objetivo de la búsqueda así como los datos obtenidos.

1.**StatsJugadoresV2.py**: Incluye la información de cada jugador y sus estadísticas en cada partido desde el 2010n a 2020. Para ello se realiza una llamada para obtener el calendario del equipo y obtener los siguientes datos.

* MP: Minutos jugados
* FG: Tiros de campos anotados
* FGA: Tiros de campo intentados
* FG%: Porcentaje de tiros de campo
* 3P: Triples anotados
* 3PA: Triples intentados
* 3P%: Porcentaje de triples
* FT: Tiros libres anotados
* FTA: Tiros libres intentados
* FT%: Porcentaje de tiros libres
* ORB: Rebotes ofensivos
* DRB: Rebotes defensivos
* TRB: Rebotes totales
* AST: Asistencias
* STL: Recuperaciones
* BLK: Tapones
* TOV: Pérdidas
* PF: Faltas personales
* PTS: Puntos
* +/-: Más menos (marcador parcial con el jugador en pista)
* TS% -- True Shooting Percentage
* eFG% -- Effective Field Goal Percentage
* 3PAr -- 3-Point Attempt Rate
* FTr -- Free Throw Attempt Rate
* ORB% -- Offensive Rebound Percentage
* DRB% -- Defensive Rebound Percentage
* TRB% -- Total Rebound Percentage
* AST% -- Assist Percentage
* STL% -- Steal Percentage
* BLK% -- Block Percentage
* TOV% -- Turnover Percentage
* USG% -- Usage Percentage
* ORtg -- Offensive Rating
* DRtg -- Defensive Rating
* BPM -- Box Plus/Minus

2. **StatsJugadoresV2\_h1.py**: Con este proceso buscamos documentar la información de cada jugador y sus estadísticas al descanso en cada partido desde 2010 a 2020. Para ello realizamos una llamada para obtener el calendario del equipo, que nos facilita la URL del boxcore de cada partido. Después organizamos el formato del dataframe para acoplarlo a nuestro modelo.

* FG: Tiros de campos anotados
* FGA: Tiros de campo intentados
* FG%: Porcentaje de tiros de campo
* 3P: Triples anotados
* 3PA: Triples intentados
* 3P%: Porcentaje de triples
* FT: Tiros libres anotados
* FTA: Tiros libres intentados
* FT%: Porcentaje de tiros libres
* ORB: Rebotes ofensivos
* DRB: Rebotes defensivos
* TRB: Rebotes totales
* AST: Asistencias
* STL: Recuperaciones
* BLK: Tapones
* TOV: Pérdidas
* PF: Faltas personales
* PTS: Puntos

3. **ResultadosClasificacionv2.py:** Creamos un dataframe con la clasificación en todas las fechas posibles desde octubre de 2009 a abril de 2020. Los datos los hemos obtenido a taves de la realización de una serie de llamadas a [www.shrpsports.com](http://www.shrpsports.com) y en los que hemos obtenido estos campos.

* Fecha (mes y día)
* Año
* Temporada
* Equipo
* Posición de conferencia
* Victorias
* Derrotas
* Porcentaje victorias / derrotas
* Diferencia con el líder de conferencia
* Victorias como local
* Derrotas como Local
* Victorias intra-división
* Derrotas intra-división
* Victorias intra-conferencia
* Derrotas intra-conferencia
* Victorias inter-conferencia
* Derrotas inter-conferencia

También hemos realizado las mismas llamadas en el df StatsjugadoresV2.py para obtener los mismos datos que obtenemos en este extractor a nivel de equipo, tanto visitante como local, obtener datos de resultaod y un flag de local ( True/False).

Para renombrar las nomenclaturas, hemos realizaod un diccionario de ambos datos obtenidos de las web ( shrsports y basketball-reference,com) para unirlos en un único dataframe. El resultado ha sido la creación de un Df que incluye todos los datos tanto de equipos visitante como del equipo local.

4.**DNP.py:** Incluye la información de qué jugadores (nombre y equipo) no participaron en los partidos durante el campeonato. Para ellos realizamos llamadas para obtener el calendario y para obtener el boxcore de cada partido, recogiendo la información del jugador que no jugó y el ID del partido. Esta información servirá para crear una variable true/false en el DataFrame que usemos como entrada del modelo de predicción que indique si el jugador estrella participó en el encuentro.

5.**Premios.py:** Recoge la información de los jugadores que han recibido votos para recibir el premio MVP (premio a mejor jugador del año) desde 2010 hasta 2020. Se registra el nombre del jugador, el número de votos en primera posición, los puntos obtenidos, los puntos máximos y el share de éstos. Esta información nos servirá para actuar como contraste al análisis de las estadísticas individuales.

6.**Distancias.py:** Mediante llamadas a [www.geodatos.net](http://www.geodatos.net/) obtenemos las distancias entre las ciudades de todos los equipos que componen la NBA. Realizamos las llamadas iterando sobre la misma lista de equipos (que se relacionan con la ciudad mediante un diccionario) dos veces, bajo la condición de que el equipo no sea el mismo. Obtenemos así la distancia de Equipo A a Equipo B en kilómetros. Del mismo modo que en el extractor DNP.py, nos servirá como una variable más para el modelo de predicción, como distancia recorrida por el equipo visitante.

7.**Evolucion.py:** Dataframe que incluye las estadísticas totales de cada equipo en cada temporada para realizar un análisis sobre la evolución en el baloncesto NBA desde 1990 hasta 2020. La fuente de estos datos es [www.basketball-reference.com](http://www.basketball-reference.com/).

* FG: Tiros de campos anotados
* FGA: Tiros de campo intentados
* FG%: Porcentaje de tiros de campo
* 3P: Triples anotados
* 3PA: Triples intentados
* 3P%: Porcentaje de triples
* FT: Tiros libres anotados
* FTA: Tiros libres intentados
* FT%: Porcentaje de tiros libres
* ORB: Rebotes ofensivos
* DRB: Rebotes defensivos
* TRB: Rebotes totales
* AST: Asistencias
* STL: Recuperaciones
* BLK: Tapones
* TOV: Pérdidas
* PF: Faltas personales
* PTS: Puntos

Una vez completado, todo el proceso de extracción de datos, hemos generado un fichero que sirva de input al modelo predictivo recogiendo toda la información.

El fichero final contiene la siguiente la información.

|  |  |
| --- | --- |
| Información | extractor |
| Clasificación antes del partido (posición, rachas, etc) | ResultadosClasificacionV2.py |
| Estadisticas al descanso | StatsJugadoresV2\_h1.py |
| Suma de los sueldos del equipo | Sueldos.py |
| AWS | StatsjugadoresV2.py |
| WP | StatsjugadoresV2.py y ResultadosClasificacionV2.py |
| Divisiones y conferencias | - |

Además, en durante este proceso, hemos eliminado toda la información duplicada o innecesaria, así como la adaptación de toda la información a un formato local /visitante.

Además de los campos obtenidos directamente de la web se realizan transformaciones para obtener los siguientes:

* Ultimos10Victorias: las victorias conseguidas en los últimos diez partidos \* jugados.
* Ultimos10Derrotas: las derrotas conseguidas en los últimos diez partidos jugados.
* Sueldo: La suma del sueldo del equipo en la temporada.
* Division: división a la que eprtenece el equipo.
* Conferencia: conferencia a la que pertenece el equipo.
* AWS\_MEDIO\_AGRUPADO: La medida AWS será explicada más adelante, una vez calculada para cada jugador, hacemos un rolling mean y un shift, para obtener la media hasta antes del partido del jugador en la temporada, posteriormente se agrupan por equipo y se realiza el join, obteniendo una valoración conjunta de los jugadores que participan en el encuentro por equipo.
* WP\_MEDIO\_AGRUPADO: Se trata de una medida análoga al AWS.

# GLOSARIO BBDD

Los primeros términos para iniciarse en las bases de datos materia deberían ser el Offensive y defensive rating, el pace o ritmo y el true shooting percentage.

1. Info general del partido:

* **Date**: Fecha
* **Year**: Año
* **Season**: Temporada
* **local\_team**: Equipo local
* **visitor\_team:**  Equipo visitante
* **Points**: puntos (equipo local)
* **Opponent\_Points**: puntos (equipo visitante)
* **Result**: Resultado
* **ID Partido**: ID partido

2. La info de la clasificacion:

* **LOCAL\_Racha**: Local victorias (número de victorias en los últimos partidos)
* **LOCAL\_Ultimos10Victorias**: Local últimas 10 victorias
* **LOCAL\_Ultimos10VDerrotas**: Local últimas 10 derrotas
* **VISITANTE\_Ultimos10Victorias**: Visitante última 10 victorias
* **VISITANTE\_Ultimos10VDerrotas**: Visitante últimas 10 derrotas
* **VISITANTE\_Racha**: Visitante victorias. ( número de victorias en los últimos partidos)
* **local\_Conf\_position**: posición en la clasificación. dentro de la conferencia ( lugar del 1 al 15)
* **local\_Win**: Local victorias
* **local\_Lose**: Local derrotas
* **local\_Percentagewl**
* **local\_Dif\_leader:**
* **local\_Home\_win**: Partidos en local ganados como local
* **local\_Home\_lose**: partidos en local perdidos como local
* **local\_Away\_win**: partidos en local ganados como visitante
* **local\_Away\_lose**: partidos perdidos como visitante
* **local\_Div\_win**: diferencia partidos ganados con el lider
* **local\_Div\_lose**: diferencia partidos perdidos con el líder
* **local\_Cnf\_win**: intracofenrecia ( este contra este)
* **local\_Cnf\_lose**
* **local\_Icf\_win** intrraconferencia ( oeste contra este
* **local\_Icf\_lose**
* **visitor\_Conf\_position**: posición equipo visitante en la clasificación dentro de la conferencia.
* **visitor\_Win**: equipo visitante victorias
* **visitor\_Lose:**  equipo visitante derrotas
* **visitor\_Percentagewl:**
* **visitor\_Dif\_leader**
* **visitor\_Home\_win**: Partidos ganados como visitante
* **visitor\_Home\_lose:**  partidos perdidos como visitante
* **visitor\_Away\_win**: partidos como visitante ganados
* **visitor\_Away\_lose**: partidos como visitante perdidos
* **visitor\_Div\_win:**  diferencia partidos ganados con el lider
* **visitor\_Div\_lose:**  diferencia partidos perdidos con el líder
* **visitor\_Cnf\_win:**  diferencia partidos ganador intraconferencia. ( este contra oeste)
* **visitor\_Cnf\_lose:**  diferencia partidos perdidos en intraconferencia ( este contra oeste)
* **visitor\_Icf\_win:**  diferencia partidos ganados en intraconferencia. ( oeste contra este)
* **visitor\_Icf\_lose:**  diferencia partidos perdidos en intraconferencua(oeste contra este)

3. Boxcore de los equipos a medio tiempo:

* **local\_fg:**  local tiros de campo
* **local\_fga** local tiros de campo intentados
* **local\_fg3:**  local tiros de triple
* **local\_fg3a**: local tiros de triple intentados
* **local\_ft:**  local tiros libres
* **local\_fta:**  local tiros libres intentados
* **local\_orb**: local rebote ofensivo
* **local\_drb**: local rebote defensivo
* **local\_trb**: suma local de rebote ofensivo y defensivo
* **local\_ast** local asistencias
* **local\_stl:**  local robos de balón
* **local\_blk**: local tapones
* **local\_tov:**  local pérdida de balones
* **local\_pf:**  local falta personal
* **local\_pts** : local puntos
* **local\_ft\_pct:**  porcentaje de puntos de equipo local
* **local\_fg\_pct:**  porcentaje de tiros de campo de equipo local
* **local\_fg3\_pct:**  porcentaje de tiros de campo equipo local
* **visitor\_fg:**  visitante tiros de campo
* **visitor\_fga**: visitante tiros de campo intentados
* **visitor\_fg3:**  visitante tiros de triple
* **visitor\_fg3a**: visitante de tiros de triple intentados
* **visitor\_ft:**  visitante tiros libres
* **visitor\_fta:**  visitante tiros libres intentados
* **visitor\_orb:**  visitante rebote ofensivo
* **visitor\_drb:**  visitante rebote defensivo
* **visitor\_trb:**  suma visitante de rebote defensivo y ofensivo
* **visitor\_ast:**  visitante asistencias
* **visitor\_stl:**  visitantes robos de balón
* **visitor\_blk:**  visitante tapones
* **visitor\_tov:**  visitante pérdida de balones
* **visitor\_pf:**  visitante falta personales
* **visitor\_pts:**  visitante puntos
* **visitor\_ft\_pct:**  porcentaje de puntos de equipo visitante
* **visitor\_fg\_pct**: porcentaje de tiros de campo de equipo visitante
* **visitor\_fg3\_pct:**  porcentaje de tiros de triples de equipo local

4. Otros.

* Sueldo local: Local sueldo total
* Sueldo visitante: Visitante sueldo total
* Local\_Division
* Local\_Conferencia
* Visitor\_Division
* Visitor\_Conferencia

# CÓMO HA EVOLUCIONADO EL JUEGO DEL BALONCESTO

<https://www.blogdebasket.com/2020/01/15/increible-asi-cambiado-zona-tiro-nba-20-anos-diferencia>

EL juego que se desarrolla actualmente en la NBA no es el mismo juego que se realizaba cuando se comenzaron a recoger datos. El análisis de datos de juego ha creado un antes y un después en el mundo del baloncesto. Tanto así que ha dado lugar a cambios importantes en la forma de jugar y las estrategias de los equipos.

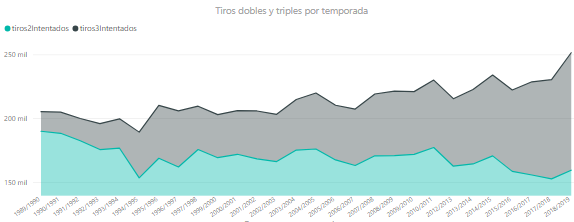
Hoy en día, cada equipo hace sus propios análisis de datos para determinar fichajes, estrategias a llevar a cabo con cada contrincante, sacar el mayor partido a cada componente del equipo, determinar minutos en cancha y partidos en los que participará cada jugador… En definitiva, toda una mina de oro para este deporte.

FORMAS EN LAS QUE SE ESTÁN OBTENIENDO DATOS

FECHA EN QUE SE COMENZÓ A ESTUDIAR LOS DATOS

La variación en la forma que se están obteniendo los datos y del año en el que comenzó a registrarse información está repercutiendo en el estilo de juego desarrollado por los equipos y en la forma de afrontar los partidos, buscando las estrategias más eficientes para conseguir el mayor número de puntos.

Con un simple vistazo a los datos de tiros por temporada a lo largo de los últimos 30 años, podemos ver que el volumen ha experimentado un gran aumento.



Esto se debe a que cada vez el juego es más rápido y por lo tanto da lugar a un mayor número de jugadas por partido. Y con ello, la oportunidad de anotar una mayor cantidad de tantos en cada encuentro.

Pero, ¿Cómo podemos conocer que el juego es más rápido? Se producen más intentos de tiro ahora que hace 20 años?

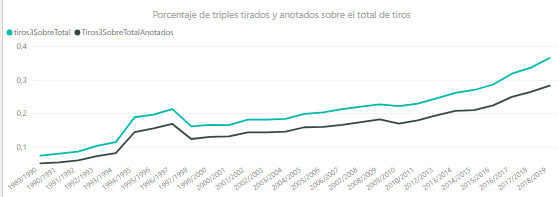
En el siguiente gráfico, comenzamos a analizar los puntos por temporada desde 1989/1990 hasta la temporada 2018/2019. La evolución se produce de forma lineal a lo largo del tiempo, teniendo un margen de crecimiento muy parecido en todos ambas metricas.



Sin embargo, el hecho de tirar más veces a canasta no es lo único que ha favorecido el aumento de puntos por temporada. Ha sido otro de los cambios tácticos importantes que ha dado la analítica de datos lo que más ha influido en este aumento.

Se trata de la importancia que se está dando a los tiros triples sobre los de 2 puntos, lo cual está invirtiendo el porcentaje de estos sobre el total de tiros, tanto en el caso de tiros intentados como en los anotados.

Como podemos ver en el siguiente gráfico, mientras en la temporada de 1989/1990, los triples sólo suponían un 7% sobre el total de tiros realizados y un 5% sobre los anotados, en la temporada de 2018/2019 estos porcentajes suponen un 37% y 28% respectivamente. Sobretodo se observa un aumento mucho más acusado en las últimas 5 temporadas, donde el incremento es de 11 y 7 puntos porcentuales respectivamente.



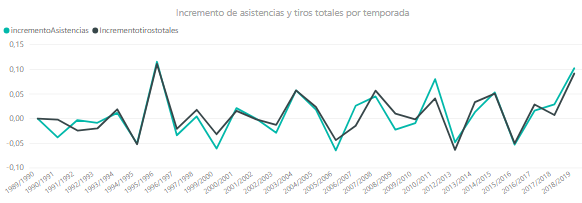
La respuesta a por qué se está llevando a cabo esta sustitución de tiros de dos puntos por tiros triples, se encuentra en la esperanza matemática que supone cada uno de ellos.

Como es evidente, el porcentaje de efectividad es superior en tiros de 2 puntos que en tiros de 3 puntos, -además este dato no ha experimentado grandes variaciones en el tiempo a pesar de que se esté tendiendo a una mayor práctica de tiros triples sobre dobles-, siendo de una media del 50% y del 35% respectivamente. Sin embargo, el hecho de anotar un triple, supone en el marcador un 50% más de puntos que un tiro doble.

Además, también se trata de una cuestión táctica. Al realizar el ataque desde fuera como es el caso de los triples, más tienen que abrirse los rivales. De esta manera resultaría más fácil atacar el aro, y por parte de la defensa, más difícil el apoyo entre los jugadores, ya que tienen que cubrir más terreno.

Una idea que podría surgir del hecho de que se esté tendiendo a un juego más basado en triples y se estén llevando a cabo jugadas más rápidas, es que el tipo de juego esté tendiendo a un modelo más individualista y menos colaborativo. Sin embargo, podemos descartar esa idea con un vistazo rápido a la evolución de los datos de tiros y asistencias en conjunto.

Como podemos observar en el siguiente gráfico, el incremento o disminución porcentual de cada temporada respecto a la anterior sigue una evolución muy similar entre ambas medidas, por lo que no podemos evidenciar un cambio en el juego hacia una tendencia más o menos individualista. Los datos de tiros totales y asistencias evolucionan en la misma dirección y en proporciones similares.



# MODELO DE EQUIPO/JUGADOR PARA SABER CUAL ES LA MEJOR VARIABLE

Con la incorporación de los datos, los equipos han logrado descubrir aquellos jugadores que por sus números no eran los mejores, pero que en el fondo desempeñaban una labor fundamental en el juego de equipo.