

PROYECTO FINAL

NBA ANALYTICS



Integrantes:
Martin Alonso
Martin Rodríguez Valiente
Fecha: 28/10/21

ÍNDICE

Introduccion:	
Descripción:	3
Presentación del equipo de trabajo:	3
Objetivo:	3
Stack Tecnológico:	3
Base de Datos:	4
Dataset:	4
Diagramas:	5
A nivel Conceptual:	5
Tablas:	5
Tabla Ranking:	5
Tabla Equipos:	6
Tabla Jugadores:	7
Tabla Game Detail:	7
Tabla Games:	8
Tablero de control:	10
Especificaciones Funcionales:	10
Alcance:	10
Usuarios:	10
Prototipado y diseño:	10
Vistas:	12
Guía de uso:	16
Especificaciones Técnicas:	17
Modelo de datos:	17
Transformaciones realizadas:	18
Medidas calculadas:	19
Segmentaciones elegidas:	20
Futuras líneas:	21

Introducción:

Descripción:

La temática por desarrollar en el proyecto será la de analizar todos los juego de las ultimas 5 (cinco) temporadas de la NBA. En el básquet como deporte, se pueden visualizar distintas facetas del juego desde las estadísticas, algunas son rígidas y otras son variables a la hora de relacionarlas con distintas estrategias/módulos de juego.

El origen de los datos estadísticos fue extraído desde la página oficial de la NBA (https://www.nba.com/stats/).

Presentación del equipo de trabajo:

El equipo de Trabajo está compuesto por:

- Martin Alonso
- Martin Rodríguez Valiente

Objetivo:

El objetivo es poder visualizar si existe un determinado patrón de juego que permita a un equipo tener mayores probabilidades de poder ser un equipo ganador y/o contendiente al título. De esta forma podría cuantificarse de mejor manera donde es mas importante realizar la inversión presupuestaria de masa salarial en el personal (jugadores) a contratar.

Stack Tecnológico:

Realizaremos la presentación en Power BI que es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas.



Base de Datos:

Dataset:

El dataset fue descargado de Kaggle (https://www.kaggle.com/nathanlauga/nba-games), de aquí se tomaron para realizar el enfoque del análisis 151.631 registros y se generaron 38 columnas que se enumeran a continuación:

- TEAM_ID
- TEAM
- SEASON
- W
- L
- Rank_ID
- PK_GAME_id
- GAME_ID
- PTS anot
- FG_PCT_anot
- FT_PCT_anot
- FG3_PCT_anot
- AST_anot
- REB_anot
- PTS_rec
- FG_PCT_rec
- FT_PCT_rec
- FG3_PCT_rec
- AST_rec
- REB_rec
- Game_detail_ID
- FGM
- FGA
- FG3M
- FG3A
- FTM
- FTA
- REB
- AST
- STL
- TOPF
- PTS
- TEAM_NICKNAME
- YEAR_FOUNDED
- ARENA_CAPACITY
- PLAYER_NAME
- PLAYER_ID

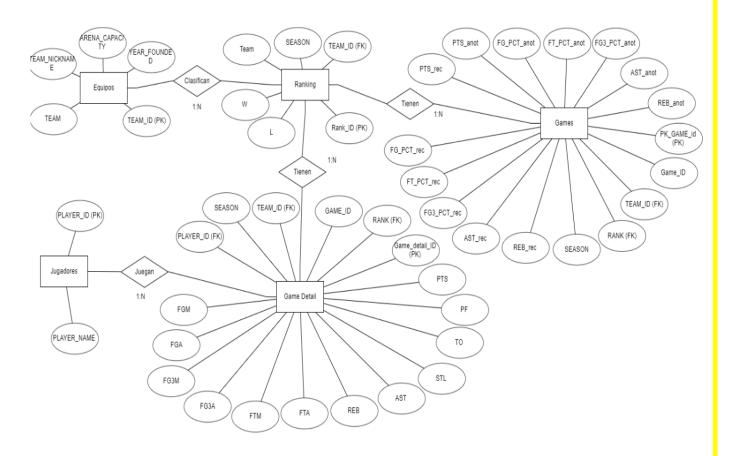
Se adjunta el dataset completo en archivo Excel:

LINK EXCEL

Diagramas:

A nivel Conceptual:

En el cuadro se puede visualizar como en la NBA interactúan nuestras 5 entidades a analizar. Donde los 30 Equipos juegan varios Partidos entre si clasificándose de menor a mayor siendo el numero 1 el que tiene mayor cantidad de partidos ganados. En los partidos se miden estadísticas a nivel equipo y a nivel jugadores.



Tablas:

Tabla Ranking:

El Ranking es una tabla de posiciones donde los equipos se clasifican en función de su récord de partidos ganados, ubicándose en primer lugar el equipo con mejor porcentaje de partidos ganados.

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
Rank_ID	Identifica unívocamente el Ranking del Equipo	Int	Si	No	No
Team_ID	Identifica unívocamente al Equipo	Int	No	Si	No
SEASON	Indica la temporada	Int	No	No	No
W	Cantidad de partidos ganados	Int	No	No	No
L	Cantidad de partidos perdidos	Int	No	No	No
TEAM	Nombre de la Ciudad del Equipo	Varchar (50)	No	No	No

Tabla Equipos:

En esta tabla se encuentran los datos completos de todos los Equipos que participan en la competencia.

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
TEAM_ID	Identifica unívocamente al Equipo	Int	Si	No	No
TEAM	Identifica a la ciudad del equipo	Varchar (50)	Si	No	No
TEAM_NICKNAME	Nombre del Equipo	Varchar (50)	No	No	No
YEAR_FOUNDED	Año de fundación del Equipo	Int	No	No	No
ARENA_CAPACITY	Capacidad de espectadores del estadio	Int	No	No	No

Tabla Jugadores:

En esta tabla tenemos los datos para poder individualizar a los jugadores a lo largo de todos los partidos de la temporada en cada una de las temporadas analizadas.

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
PLAYER_ID	Identifica unívocamente al Jugador	Int	Si	No	No
PLAYER_NAME	Nombre del Jugador	Varchar (50)	No	No	No

Tabla Game Detail:

En esta tabla tenemos todas las estadísticas a analizar partido a partido de cada una de las temporadas de todos los equipos de la competencia

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
GAME_DETAIL_I D	Identifica unívocamente a la estadística del jugador en el partido	Varchar (50)	Si	No	No
GAME_ID	Identifica el partido	Int	No	Si	No
TEAM_ID	Identifica el equipo	Int	No	Si	No
SEASON	Temporada	Int	No	No	No
PLAYER_ID	Identifica al jugador	Int	No	Si	No
REB	Rebotes realizados por el jugador en el partido	Int	No	No	Si
AST	Asistencias realizados por el jugador en el partido	Int	No	No	Si

STL	Robos realizados por el jugador en el partido	Int	No	No	Si
FGM	Tiros de cancha intentados	Int	No	No	Si
FGA	Tiros de cancha acertados	Int	No	No	Si
FG3M	Tiros de tres puntos intentados	Int	No	No	Si
FG3A	Tiros de tres puntos acertados	Int	No	No	Si
FTM	Tiros libres intentados	Int	No	No	Si
FTA	Tiros libre acertados	Int	No	No	Si
ТО	Perdidas realizados por el jugador en el partido	Int	No	No	Si
PTS	Puntos realizados por el jugador en el partido	Int	No	No	Si
RANK	Indica el ranking en la temporada correspondiente	Int	No	Si	No

Tabla Games:

En la tabla partidos se registran los resultados que analizaremos en relación con el equipo ganador y al perdedor, como así también las estadísticas a favor y en contra de cada equipo.

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
PK_GAME_id	Identifica unívocamente a un partido de un equipo	Varchar (50)	Si	No	No
Game_ID	Identifica a un partido de la temporada	Int	No	No	No

TEAM_ID	Identifica al Equipo	Int	No	Si	No
PTS_anot	Puntos anotados por el Equipo	Int	No	No	Si
FG_PCT_anot	Porcentaje de tiros de campo anotados por el equipo	Float	No	No	Si
FT_PCT_anot	Porcentaje de tiros libres anotados por el equipo	Float	No	No	Si
FG3_PCT_anot	Porcentaje de tiros de tres puntos anotados por el equipo	Float	No	No	Si
AST_anot	Asistencias realizadas por el Equipo	Int	No	No	Si
REB_anot	Rebotes realizados por el Equipo	Int	No	No	Si
PTS_rec	Puntos recibidos por el Equipo	Int	No	No	Si
FG_PCT_rec	Porcentaje de tiros de campo recibidos por el equipo	Float	No	No	Si
FT_PCT_rec	Porcentaje de tiros libres recibidos por el equipo	Float	No	No	Si
FG3_PCT_rec	Porcentaje de tiros de tres puntos recibidos por el equipo	Float	No	No	Si
AST_rec	Asistencias recibidas por el Equipo	Int	No	No	Si
REB_rec	Rebotes recibidos por el Equipo	Int	No	No	Si
RANK	Indica el ranking en la temporada correspondiente	Int	No	Si	No

SEASON	Indica la temporada	Int	No	No	No

Tablero de control:

Especificaciones Funcionales:

Alcance:

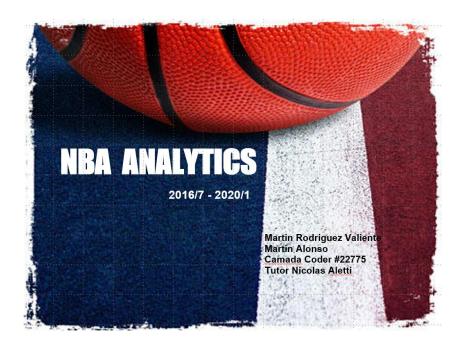
En el tablero se podrán visualizar estadísticas relacionadas a los partidos que se realizaron en el marco de la National Basketball Association (NBA) entre los años 2016 y 2020. Se podrá ver los resultados en cuanto a partidos ganados y perdidos de cada uno de los 30 equipos que componen la liga. A su vez, se analiza el desempeño en el aspecto ofensivo y defensivo a nivel equipo y jugador, utilizando distintas estadísticas de juego por partido como: puntos, asistencias, rebotes y la eficacia de tiros de campo, de dos puntos y tres puntos.

Usuarios:

El tablero realizado está dirigido a usuarios involucrados en la administración de equipos de baloncesto. Presenta estadísticas de los equipos y jugadores, que participaron en las últimas cinco temporadas, y los resultados que obtuvieron. De esta forma, busca analizar qué aspectos son los más relevantes a la hora de contratar nuevo talento a un equipo.

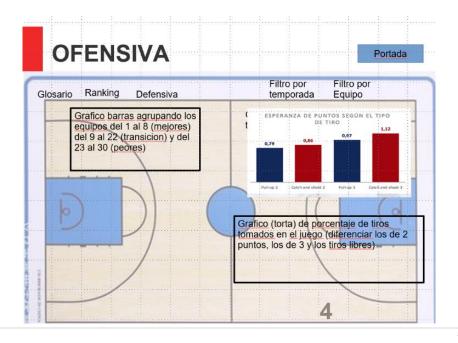
Prototipado y diseño:

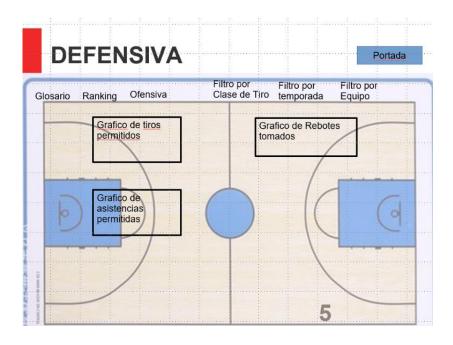
A continuación, se presentan imágenes del prototipado original de las solapas que contiene el tablero:











En el desarrollo del prototipado a su versión final se seleccionaron los tipos de gráficos adecuados a la información que se quiere transmitir, se definieron los filtros a utilizar en todas las solapas, y se agregaron indicadores de distintas estadísticas relevantes del aspecto ofensivo y defensivo de los equipos.

Vistas:

El tablero cuenta con cinco solapas: Portada, Glosario, Ranking, Ofensiva y Defensiva.

Portada:

Indica el nombre del proyecto, una descripción de los datos utilizados, y un set de botones de navegación del tablero. Tanto esta solapa como las restantes cuentan en la esquina superior izquierda con el logo de la liga a la que los datos hacen referencia.



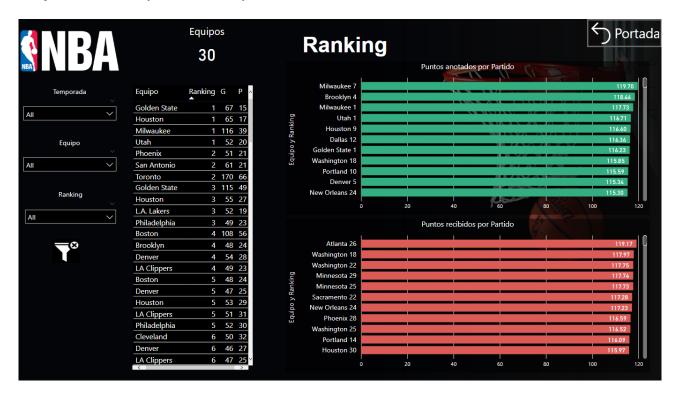
Glosario:

El glosario muestra una gráfica que identifica los distintos tipos de tiros al aro, diferenciación que luego aparecerá en el tablero, de forma de que cualquier persona pueda interpretarlos y contrastar entre estos. A su vez, indica la definición de asistencia y rebote. Tanto esta solapa como las siguientes cuentan en la esquina superior derecha con un botón de navegación hacia la portada.



Ranking:

La siguiente solapa muestra la cantidad total de equipos que participan en la liga, la posición de cada equipo en el ranking, con sus respectivos partidos ganados y perdidos, al seleccionar una temporada se mostrarán los equipos de la posición 1 a la 30 de la misma. Se muestran dos gráficos que representan los puntos anotados y recibidos por partido de cada equipo y su correspondiente ranking. Tanto esta solapa como las siguientes cuentan con tres filtros, Temporada: se puede elegir una temporada del intervalo 2016-2020, Equipo: se puede seleccionar uno o múltiples equipos de los 30 participantes y Ranking: se puede elegir entre un rango de equipos según su posición (Mejores 5, Entre 6 y 15, Entre 16 y 25, Peores 5)



Ofensiva:

La misma cuenta con dos indicadores: Puntos anotados por partido y Asistencias realizadas por partido.

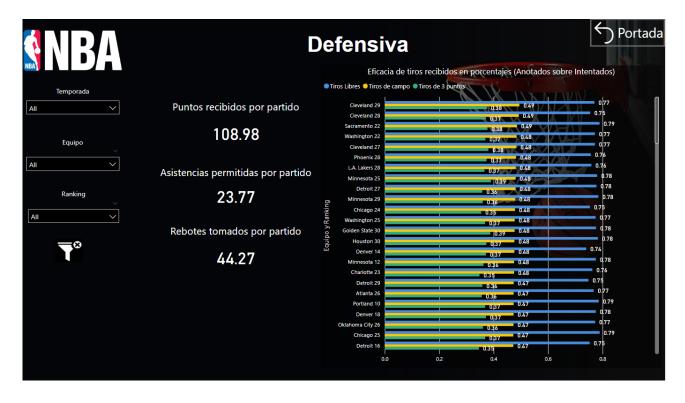
Una tabla con el promedio de puntos por partido de cada jugador, junto a un filtro que modifica la tabla, y que se compone de un parámetro que se puede editar para que refleje solamente a los jugadores con un promedio de puntos por partido mayor al valor del parámetro.

Un gráfico de la proporción de tiros tomados, entre tiros de 2 y 3 puntos. Junto a un gráfico mostrando la eficacia de los mismos, desagregado por cada equipo y por tiros de campo, de 3 puntos y tiros libres.



Defensiva:

La última solapa cuenta con tres indicadores relacionados al aspecto defensivo de un equipo: Puntos recibidos, asistencias permitidas y rebotes tomados por partido. Junto a un gráfico con la eficacia de los tiros recibidos, es decir los tiros en contra a su canasto mostrando cuanta eficacia permite en los tiros de su rival.



Guía de uso:

Para navegar entre las solapas se crearon botones de interacción, que se pueden acceder haciendo ctrl + click izq sobre ellos.

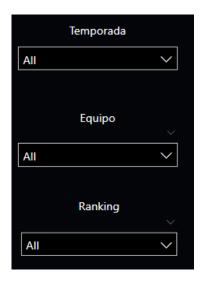
Se creó en la portada un botón para cada solapa:



En cada solapa se encuentra un botón para poder regresar a la portada:



Se agregaron filtros en cada solapa, que interaccionan entre sí, por lo que si, por ejemplo, se modifica un filtro en la solapa de Ranking, el mismo también se modifica en las solapas Ofensiva y Defensiva. Se puede filtrar por Temporada, Equipo y Ranking.



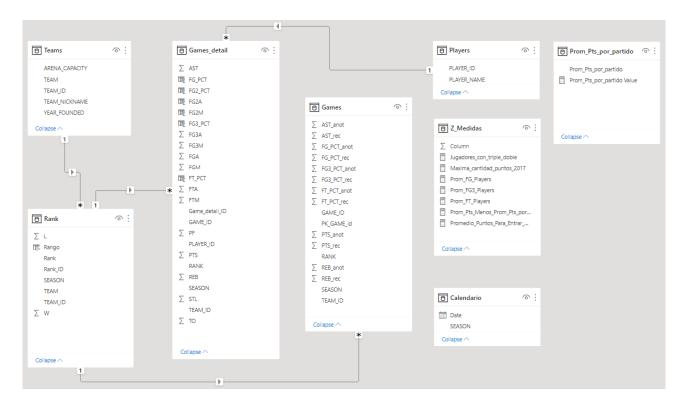
A su vez, se agregó un botón para borrar los filtros que se hayan realizado, de esta forma al acceder a ese botón se volverá la vista a la forma original:



Especificaciones Técnicas:

Modelo de datos:

Diagrama de entidad-relación que refleja la herramienta Power BI, donde se muestra la tabla creada Calendario, y las transformaciones realizadas, tanto columnas creadas, como las medidas creadas en Z_Medidas y el parámetro creado en Prom_Pts_por_partido.



La tabla Calendario se compone de dos columnas y se utilizó para realizar la medida Maxima_cantidad_puntos_2017, que se detalla en la siguiente sección.

Tabla Calendario:

Columnas	Descripción	Tipo de Dato	Es PK	Es FK	Null
SEASON	Identifica al año de la temporada	Int	Si	No	No

	Identifica al primer día del año de la				
Date	temporada	date	No	No	No

Transformaciones realizadas:

En la tabla Rank, se crearon las columnas:

- Rank
- Rango.

La columna Rank se crea mediante un ejemplo, para extraer, a partir de la columna Rank_ID, la posición exacta en la que se ubica cada equipo por temporada.

La columna Rango se crea para clasificar a los equipos según su posición en el ranking, utilizando la siguiente formula:

```
Rango = IF('Rank'[Rank] <= 5, "Mejores 5", IF(AND( 'Rank'[Rank] >= 6, 'Rank'[Rank] <= 15) , "Entre 6 y 15", IF(AND( 'Rank'[Rank] >= 16, 'Rank'[Rank] <= 25) , "Entre 16 y 25", "Peores 5") ) ) ) \\
```

En la tabla Games_detail, se crearon las columnas:

- FG_PCT
- FG3 PCT
- FT_PCT
- FG2A
- FG2M
- FG2 PCT

La columna FG_PCT se crea para calcular el porcentaje de tiros de campo anotados sobre tiros de campo realizados en un partido, por cada jugador. Con la siguiente formula:

```
FG PCT = IFERROR(Games detail[FGM] / Games detail[FGA], 0)
```

La columna FG3_PCT se crea para calcular el porcentaje de tiros de tres puntos anotados sobre tiros de tres puntos realizados en un partido, por cada jugador. Con la siguiente formula:

```
FG3_PCT = IFERROR(Games_detail[FG3M]/Games_detail[FG3A],0)
```

La columna FT_PCT se crea para calcular el porcentaje de tiros libres anotados sobre tiros libres realizados en un partido, por cada jugador. Con la siguiente formula:

```
FT_PCT = IFERROR(Games_detail[FTM] / Games_detail[FTA],0)
```

La columna FG2A se crea para calcular la cantidad de tiros de 2 puntos realizados¹. Con la siguiente formula:

```
FG2A = Games\_detail[FGA] - Games\_detail[FG3A]
```

_

¹ Al ser (tiros de campo = tiros de 2 puntos + tiros de 3 puntos), se realiza la diferencia entre tiros de campo y tiros de 3 puntos para obtener la cantidad de tiros de 2 puntos realizados.

La columna FG2M se crea para calcular la cantidad de tiros de 2 puntos anotados. Con la siguiente formula:

```
FG2M = Games\_detail[FGM] - Games\_detail[FG3M]
```

La columna FG2_PCT se crea para calcular el porcentaje de tiros de 2 puntos anotados sobre tiros de 2 puntos realizados en un partido, por cada jugador. Con la siguiente formula: FG2_PCT = IFERROR(Games_detail[FG2M] / Games_detail[FG2A],0)

Se creó la tabla: Calendario que contiene dos columnas. Una columna SEASON, que se obtuvo duplicando la columna SEASON de la tabla Rank y dejando los valores únicos: [2016,2017,2018,2019,2020]. Y una columna Date, que se obtuvo a partir de un ejemplo, para obtener el primer día de cada año de la columna SEASON.

Se creó la tabla: Z_Medidas que contiene las medidas realizadas en este proyecto.

Medidas calculadas:

- Prom_FG_Players
- Prom_FG3_Players
- Prom_FT_Players
- Prom_Pts_Menos_Prom_Pts_por_partido_Value
- Promedio_Puntos_Para_Entrar_A_Playoffs
- Maxima_cantidad_puntos_2017
- Jugadores_con_triple_doble

Prom_FG_Players: Calcular el promedio de porcentaje de tiros de campo de los jugadores. Formula: Prom_FG_Players = AVERAGE(Games_detail[FG_PCT])

Prom_FG3_Players: Calcular el promedio de porcentaje de tiros de tres puntos de los jugadores. Formula: Prom_FG3_Players = AVERAGE(Games_detail[FG3_PCT])

Prom_FT_Players: Calcular el promedio de porcentaje de tiros libres de los jugadores. Formula: Prom FT Players = AVERAGE(Games detail[FT PCT])

Prom_Pts_Menos_Prom_Pts_por_partido_Value: Calcula la diferencia entre el promedio de puntos por partido y un parámetro de puntos por partido. Para esta medida se creó el <u>Parámetro</u> Prom_Pts_por_partido: de forma de poder interactuar con el gráfico de tabla, que refleja el promedio de puntos por partido de cada jugador.

```
Formula: Prom_Pts_Menos_Prom_Pts_por_partido_Value = AVERAGE('Games_detail'[PTS]) - [Prom_Pts_por_partido Value]
```

Para el parámetro Prom Pts por partido, se utiliza un intervalo de números enteros desde 1 a 34.

Promedio_Puntos_Para_Entrar_A_Playoffs: Calcula el promedio de puntos por partido anotados por los equipos que lograron tener más victorias en una temporada que el promedio histórico de victorias necesarias para entrar a los Playoffs²

² Se toma el promedio de victorias de los equipos que se ubicaron en la posición 16 del ranking, ya que es el último equipo que clasifica a los playoffs.

```
Formula: Promedio_Puntos_Para_Entrar_A_Playoffs =
VAR Promedio_Victorias_Para_Entrada_A_Playoffs = INT(CALCULATE(AVERAGE('Rank'[W]),'Rank'[Rank] ==
16))
return CALCULATE(AVERAGE(Games[PTS_anot]),'Rank'[W] > Promedio_Victorias_Para_Entrada_A_Playoffs
)
```

Maxima_cantidad_puntos_2017: Calcula la cantidad máxima de puntos que fueron anotados en un partido en la temporada 2017.

```
Formula: Maxima_cantidad_puntos_2017 =
var Anio2020 = LASTDATE(Calendario[Date])
var Anio2017 = DATEADD(Anio2020,-3,year)
var max_puntos_2017 = CALCULATE(MAX(Games[PTS_anot]), Games[SEASON] == YEAR(Anio2017) )
return max_puntos_2017
```

Jugadores_con_triple_doble³: Calcula la cantidad de jugadores, por temporada, que lograron al menos un triple doble.

```
Formula: Jugadores_con_triple_doble =
var Jugadores_con_triple_doble2016 = CALCULATE(DISTINCTCOUNT(Games_detail[PLAYER_ID]) ,
Games_detail[PTS] > 9 , Games_detail[AST] > 9 , Games_detail[REB] > 9 , Games_detail[SEASON] ==
2016)
var Jugadores con triple doble2017 = CALCULATE(DISTINCTCOUNT(Games detail[PLAYER ID]) ,
Games_detail[PTS] > 9 , Games_detail[AST] > 9 , Games_detail[REB] > 9 , Games_detail[SEASON] ==
var Jugadores con triple doble2018 = CALCULATE(DISTINCTCOUNT(Games detail[PLAYER ID]) ,
Games_detail[PTS] > 9 , Games_detail[AST] > 9 , Games_detail[REB] > 9 , Games_detail[SEASON] ==
var Jugadores_con_triple_doble2019 = CALCULATE(DISTINCTCOUNT(Games_detail[PLAYER_ID]) ,
Games detail[PTS] > 9 , Games detail[AST] > 9 , Games detail[REB] > 9 , Games detail[SEASON] ==
2019)
var Jugadores_con_triple_doble2020 = CALCULATE(DISTINCTCOUNT(Games_detail[PLAYER_ID]) ,
\label{lem:continuous} {\tt Games\_detail[PTS]} \ > \ 9 \ \ , \ {\tt Games\_detail[REB]} \ > \ 9 \ \ , \ {\tt Games\_detail[SEASON]} \ == \ \ .
2020)
return COMBINEVALUES(" -
",Jugadores_con_triple_doble2016,Jugadores_con_triple_doble2017,Jugadores_con_triple_doble2018,Ju
gadores_con_triple_doble2019, Jugadores_con_triple_doble2020)
```

Segmentaciones elegidas:

Como segmentación se eligieron los siguientes campos:

```
"Rank" [TEAM] De forma de filtrar por Equipo.
"Rank" [SEASON] De forma de filtrar por Temporada.
"Rank" [Rango] De forma de filtrar por resultado en el ranking.
```

³ Triple doble es un término popular en el baloncesto que hace referencia a lograr dobles dígitos (un mínimo de 10) en tres categorías (puntos, rebotes y asistencias) en un partido.

Futuras líneas:

Dentro de las conclusiones que se extraen del análisis del tablero, se encuentran la poca relevancia de los rebotes en cuanto a la posición en el ranking, en contraste se observa una mayor diferencia en las asistencias por partido, donde los que resultan en el top 5 del ranking tienen como mínimo un promedio de 25 asistencias por partido.

Un cambio interesante que ocurrió en este intervalo es el aumento de la proporción de tiros de 3 puntos año a año, pasando de un promedio de liga de 31,7% en 2016 a 39,3% en 2020, donde en los equipos top 5 llega a un 40,5%.

También se aprecia un aumento del promedio de puntos que se anotan por partido, pasando de 105 puntos en 2016 a 112 puntos en 2020, que podría ser una consecuencia del cambio de tendencia en los tiros de larga distancia. A su vez, se aprecia que los equipos top 5 no sólo tienen un jugador estrella que lidera las anotaciones, sino que cuentan con mínimo tres o cuatro jugadores por encima de los 15 puntos promedio por partido.

Del análisis se desprende la necesidad de tener jugadores que realmente jueguen en equipo y generen un mayor promedio de asistencias por partido. Así mismo, la tendencia en aumento de los tiros de tres puntos no parece terminar, esto se debe tener en cuenta a la hora de contratar jugadores, que cuenten con buen porcentaje de este tiro, o guiar los esfuerzos en mejorar al actual talento del equipo en este aspecto.