



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

ADE

Facultad de Administración  
y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Identificación de las estrategias de los equipos de la NBA y  
su influencia en la clasificación durante una temporada  
regular.

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Prado Raposo, Sergio

Tutor/a: Martínez Gómez, Mónica

CURSO ACADÉMICO: 2022/2023

# **IDENTIFICACIÓN DE LAS ESTRATEGIAS DE LOS EQUIPOS DE LA NBA Y SU INFLUENCIA EN LA CLASIFICACIÓN DURANTE UNA TEMPORADA REGULAR**

**Sergio Prado Raposo**

**Tutor: Mónica Martínez Gómez**

Trabajo Fin de Grado presentado en la Facultad de  
Administración y Dirección de Empresas de la  
Universitat Politècnica de València, para la obtención del  
grado en Administración y Dirección de Empresa

Curso 2022-23

Valencia, 6 de julio de 2023



## Agradecimientos

*Este proyecto va dedicado a mi familia, que me ha brindado la oportunidad de poder estudiar esta titulación, sin ellos no hubiera sido posible llegar hasta aquí. Pero en especial, agradecer a mi madre por estar siempre ahí y darme la ayuda y el apoyo necesario para poder superar cada uno de los obstáculos que han aparecido por el camino.*

*Agradecer también a mi tutora Mónica por darme la oportunidad de trabajar en este proyecto. Gracias por compartir tus conocimientos conmigo, por ayudarme en todo momento a conseguir los objetivos y, sobre todo, por hacerlo siempre de una forma tan amable. Gracias a tu apoyo ha sido más fácil sacar el proyecto adelante.*

*Por último, agradecer a mis amigos por estar siempre dispuestos a ayudarme y por animarme en cada uno de los peores momentos. Y, por supuesto, a mi familia Erasmus, gracias a ellos he podido vivir la mejor experiencia de mi vida.*



## Resumen

El análisis estadístico está adquiriendo cada vez un peso mayor en el deporte. Estos análisis ayudan a recabar todo tipo de información del juego, tanto del equipo en global como de cada uno de los jugadores. En el presente trabajo se han recopilado los datos estadísticos de una temporada regular de cada uno de los equipos de la NBA. Estas estadísticas se les ha tratado por diversos métodos estadísticos, como es el Análisis de Componentes Principales (PCA) y el análisis de clúster, donde se ha podido identificar cada una de las estrategias y definir la influencia que ha tenido en su clasificación final en la tabla.

**Palabras clave:** información del juego, estrategias, influencia.

## Resum

L'anàlisi estadístic està adquirint cada vegada un pes major en l'esport. Aquestes anàlisis ajuden a recaptar tot tipus d'informació del joc, tant de l'equip en global com de cadascun dels jugadors. En el present treball s'han recopilat les dades estadístiques d'una temporada regular de cadascun dels equips de la NBA. Aquestes estadístiques se'ls ha tractat per diversos mètodes estadístics com és el PCA i l'anàlisi de clúster, on s'ha pogut identificar cadascuna de les estratègies i definir la influència que ha tingut en la seua classificació final en la taula.

**Paraules clau:** informació del joc, estratègies, influència.

## Abstract

Statistical analysis is becoming increasingly important in sports. These analyses help to gather all kinds of information about the game, both team as a whole and of each individual player. In the present work, the statistical data of a regular season of each of the NBA teams have been compiled. These statistics have been treated by various statistical methods such as PCA and cluster analysis, where it has been possible to identify each of the strategies and define the influence it has had on its final ranking in the table.

**Keywords:** information about the game, strategies, influence.



## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Gráfico de scores (a) y gráfico de loadings (b).....	15
<b>Figura 2:</b> Métodos de Agrupación.....	16
<b>Figura 3:</b> Método no jerárquico.....	16
<b>Figura 4:</b> Método Jerárquico.....	17
<b>Figura 5:</b> Dendograma.....	17
<b>Figura 6:</b> Análisis discriminante lineal.....	20
<b>Figura 7:</b> Puntos como local (T.21-22).....	36
<b>Figura 8:</b> Puntos como visitante (T.21-22).....	36
<b>Figura 9:</b> Diferencia como local (T.21-22).....	36
<b>Figura 10:</b> Diferencia como visitante (T.21-22).....	37
<b>Figura 11:</b> T2 como local (T.21-22).....	37
<b>Figura 12:</b> T2 como visitante (T.21-22).....	38
<b>Figura 13:</b> T3 como local (T.21-22).....	38
<b>Figura 14:</b> T3 como visitante (T.21-22).....	38
<b>Figura 15:</b> TC como local (T.21-22).....	39
<b>Figura 16:</b> TC como visitante (T.21-22).....	39
<b>Figura 17:</b> TL como local (T.21-22).....	39
<b>Figura 18:</b> TL como visitante (T.21-22).....	40
<b>Figura 19:</b> Rebotes como local (T.21-22).....	40
<b>Figura 20:</b> Rebotes como visitante (T.21-22).....	40
<b>Figura 21:</b> AST, ROB, TAP, PÉR y FP como local (T.21-22).....	41
<b>Figura 22:</b> AST, ROB, TAP, PÉR y FP como visitante (T.21-22).....	41
<b>Figura 23:</b> Resultado test Kolmogorov-Smirnov.....	43
<b>Figura 24:</b> Gráfica correlaciones.....	45
<b>Figura 25:</b> Resultados PCA.....	46
<b>Figura 26:</b> Resultado método Kaiser.....	47
<b>Figura 27:</b> Loadings obtenidos.....	48
<b>Figura 28:</b> Varianza explicada de cada componente.....	49
<b>Figura 29:</b> Gráfico componente RC1.....	50
<b>Figura 30:</b> Gráfico componente RC2.....	51
<b>Figura 31:</b> Gráfico componente RC3.....	53
<b>Figura 32:</b> Gráfico componente RC4.....	54
<b>Figura 33:</b> Resultados PCA.....	55
<b>Figura 34:</b> Resultado método Kaiser.....	55
<b>Figura 35:</b> Loadings obtenidos.....	56
<b>Figura 36:</b> Varianza explicada de cada componente.....	56
<b>Figura 37:</b> Gráfico componente RC1.....	58
<b>Figura 38:</b> Gráfico componente RC2.....	59
<b>Figura 39:</b> Gráfico componente RC3.....	60
<b>Figura 40:</b> Gráfico componente RC4.....	61
<b>Figura 41:</b> Gráfica componente RC5.....	61
<b>Figura 42:</b> Dendograma.....	63
<b>Figura 43:</b> Método del "codo".....	64



<b>Figura 44:</b> Prueba de chi-cuadrado .....	66
<b>Figura 45:</b> Dendograma .....	67
<b>Figura 46:</b> Método del "codo" .....	68
<b>Figura 47:</b> Prueba de chi-cuadrado .....	70
<b>Figura 48:</b> Curva ROC .....	71
<b>Figura 49:</b> Peso de las variables en las predicciones .....	72
<b>Figura 50:</b> Curva ROC .....	73
<b>Figura 51:</b> Peso de las variables en las predicciones .....	74
<b>Figura 52:</b> Estadística filtrada por eficiencia ofensiva .....	76
<b>Figura 53:</b> Clasificación Conferencia Este .....	76
<b>Figura 54:</b> Estadística filtrada por eficiencia ofensiva .....	77
<b>Figura 55:</b> Clasificación Conferencia Oeste .....	77
<b>Figura 56:</b> Script RStudio. Análisis Descriptivo .....	82
<b>Figura 57:</b> Script RStudio. Análisis Descriptivo (1) .....	82
<b>Figura 58:</b> Código RStudio Test Kolmogorov-Smirnov .....	82
<b>Figura 59:</b> Script cálculo correlación .....	82
<b>Figura 60:</b> Análisis PCA .....	83
<b>Figura 61:</b> Criterio de Kaiser .....	83
<b>Figura 62:</b> Obtención cargas factoriales .....	83
<b>Figura 63:</b> Método de Ward .....	83
<b>Figura 64:</b> Método del codo .....	84
<b>Figura 65:</b> Método del codo (1) .....	84
<b>Figura 66:</b> Método del codo (2) .....	85
<b>Figura 67:</b> Proporción V/D en cada clúster .....	85
<b>Figura 68:</b> Análisis discriminante .....	85
<b>Figura 69:</b> Análisis discriminante (2) .....	86
<b>Figura 70:</b> Análisis discriminante (3) .....	86



## ÍNDICE DE TABLAS

<b>Tabla 1:</b> Variable incluidas en la base de datos NBA.	22
<b>Tabla 2:</b> Variable incluidas en la base de datos NBA (1).	23
<b>Tabla 3:</b> Variables incluidas en la base de datos NBA (2).	25
<b>Tabla 4:</b> Variables incluidas en la base de datos NBA (3).	27
<b>Tabla 5:</b> Variables incluidas en la base de datos NBA (4).	29
<b>Tabla 6:</b> Variables incluidas en la base de datos NBA (5).	31
<b>Tabla 7:</b> Variables incluidas en la base de datos NBA (6).	31
<b>Tabla 8:</b> Variables Base de Datos.	32
<b>Tabla 9:</b> Variables seleccionadas.	35
<b>Tabla 10:</b> Tabla con el primer análisis descriptivo.	42
<b>Tabla 11:</b> Valores método Kaiser.	47
<b>Tabla 12:</b> Estrategias seguidas cada componente.	54
<b>Tabla 13:</b> Cargas Factoriales de las Componentes.	55
<b>Tabla 14:</b> Estrategias seguidas cada componente.	62
<b>Tabla 15:</b> Balanceo de los clústeres.	64
<b>Tabla 16:</b> Scores obtenidos.	65
<b>Tabla 17:</b> Proporción V/D obtenidos según estrategia.	65
<b>Tabla 18:</b> Balanceo de los clústeres.	68
<b>Tabla 19:</b> Scores obtenidos.	69
<b>Tabla 20:</b> Proporción V/D obtenidos según estrategia.	69
<b>Tabla 21:</b> Matriz de confusión.	71
<b>Tabla 22:</b> Estrategia definida.	73
<b>Tabla 23:</b> Matriz de confusión.	74



## Lista de Acrónimos

BBDD - Base de Datos

NBA – National Basketball Association

IDE – Entorno de Desarrollo Integrado

PCA – Análisis de Componentes Principales

ESPN - Entertainment and Sports Programming Network

TCI – Tiros de Campo Intentados

T2I – Tiros de 2 puntos Intentados

T3I – Tiros de 3 puntos Intentados

TCA – Tiros de Campo Anotados

ROC - Receiver Operating Characteristic

T. 21-22 – Temporada 2021-2022



AGRADECIMIENTOS	2
RESUMEN	3
ÍNDICE DE FIGURAS	4
ÍNDICE DE TABLAS	6
LISTA DE ACRÓNIMOS	7
CAPÍTULO 1. INTRODUCCIÓN	10
<b>1.1. Motivación</b>	<b>10</b>
<b>1.2. Objetivos</b>	<b>11</b>
<b>1.3. Metodología</b>	<b>11</b>
1.3.1. Software utilizado	12
1.3.2. Aprendizaje no supervisado	13
1.3.2.1. Análisis de Componentes Principales	13
1.3.2.2. Análisis Clúster	15
1.3.3. Aprendizaje supervisado	18
1.3.3.1. Análisis Discriminante	19
CAPÍTULO 2. ANÁLISIS DE LA BASE DE DATOS	21
<b>2.1. Base de datos</b>	<b>21</b>
2.1.1. Descripción de la base de datos	21
2.1.2. Selección de variables:	32
2.1.2.1. Análisis descriptivo inicial de las estadísticas de los equipos mediante gráficos	35
2.1.2.2. Análisis descriptivo de las variables. Estudio de la normalidad	41
CAPÍTULO 3. ANÁLISIS DE RESULTADOS	45
<b>3.1. Estudio de la estrategia de los equipos de la NBA</b>	<b>45</b>
3.1.1. Análisis de componentes principales	46
3.1.1.1. Equipos como local	46
3.1.1.2. Equipo como visitante	54
3.1.2. Análisis Clúster	62
3.1.2.1. Análisis clúster jerárquico	62
3.1.2.1.1. Equipos como local	62
3.1.2.1.2. Equipos como visitante	67
3.1.3. Análisis discriminante	70
1.3.2.1 Análisis discriminante equipos como local	70
	8



1.3.2.2 Análisis discriminante equipos como visitante	73
<b>3.2. Influencia en la clasificación de las estrategias</b>	<b>75</b>
<b>CAPÍTULO 4. CONCLUSIONES Y LÍNEAS FUTURAS.</b>	<b>78</b>
<b>4.1. Conclusiones</b>	<b>78</b>
<b>4.2. Líneas Futuras</b>	<b>78</b>
<b>BIBLIOGRAFÍA</b>	<b>79</b>
<b>ANEXO</b>	<b>82</b>
<b>Scripts de RStudio</b>	<b>82</b>

## Capítulo 1. Introducción

Actualmente, el análisis de los datos se ha convertido en una técnica esencial para cualquier ámbito profesional. En concreto, dentro del ámbito deportivo, se manejan inmensos conjuntos de datos de información que, gracias al Big Data, puede ayudar a mejorar la toma de decisiones, tanto a nivel de rendimiento como a nivel de gestión de fichajes, economía...

Durante cualquier práctica deportiva se producen miles de datos, que desde el punto de vista humano es imposible interpretar, aquí es donde el Big Data se está convirtiendo en una herramienta indispensable para poder recapitular todos estos datos, analizarlos y obtener unas conclusiones que ayudan a la obtención de resultados óptimos de los equipos y, a su vez, contribuyan a mejorar los rendimientos de cada uno de los deportistas profesionales.

Uno de los estudios pioneros de esta tecnología fue el desarrollado por Bill James, aspirante a articulista deportivo de Kansas, recopiló datos sobre el juego de cada jugador de las Grandes Ligas de béisbol, en función de los datos estadísticos que se disponían. Estos datos, eran totalmente objetivos, no tenía ningún peso subjetivo, y aportaban gran valor a los jugadores.

En las últimas décadas, técnicas de Big Data, Inteligencia Artificial y Machine Learning se han vuelto indispensables en los clubes de élites deportivos, para analizar los grandes volúmenes de datos que se producen que ayuden a definir nuevas estrategias que les generen ventajas competitivas que contribuyan al éxito de juego.

### 1.1. Motivación

Hace ya muchos años en los que comencé mi andadura en el deporte, y tras muchos pasos por el fútbol, tenis, balonmano... finalmente me decanté por el que ahora sigue siendo el deporte que practico, el baloncesto.

Cuando comencé a jugar y ver baloncesto, me apasionaba seguir las estadísticas de los jugadores, y a su vez la de los equipos. En ellas se podía ver muchas veces la influencia que había tenido un jugador dentro del propio equipo, tanto de forma positiva como de forma negativa. A su vez, muchos equipos se decantaban por un estilo de juego mucho más directo, a través de jugadas rápidas de contrataque, otros equipos jugaban de forma más pausada buscando el lanzamiento de tres... en conclusión, las estadísticas desprendían el estilo de juego característico de cada uno de los equipos.

Además, tras llevar varios años jugando, me gustó la idea de ser entrenador de baloncesto. Dentro de esta faceta, es mucho más importante el análisis del equipo rival y reconocer las cualidades del equipo contrario para poder contrarrestarlas.

Es por esto por lo que, una vez terminada la carrera de ADE y ver los distintos enfoques de la titulación, el ámbito de la estadística es uno de los que me ha parecido más interesante, ya que, a través de distintos datos, puedes obtener unas conclusiones muy interesantes y concluyentes acerca de cualquier ámbito profesional.

Por ende, pensé en la posibilidad de integrar el ámbito de la estadística con el baloncesto en mi trabajo final de grado. Esto conllevaría aplicar técnicas avanzadas de datos con la finalidad de



desarrollar un estudio de las estadísticas de los equipos para poder determinar estrategias de juego y ayudar a los profesionales del baloncesto a desarrollar la planificación deportiva y poder aplicarlo a cualquier liga profesional, semiprofesional o amateur con la idea de que pueda ayudar, tanto a mí, como a otros muchos entrenadores en el análisis del equipo contrario, para intentar explotar tanto las propias virtudes de nuestro equipo, como aprovechar los puntos débiles del rival, para alcanzar los objetivos marcados por cada equipo.

Además, cada vez es más habitual ver en los distintos partidos como los narradores, comentaristas y analistas deportivos se apoyan en datos estadísticos. Estos datos, algunas veces tratan sobre la posibilidad que tiene un equipo de ganar en función del resultado y el minuto en que se encuentran, otras veces muestran la posibilidad que tiene un jugador de fútbol de lanzar un penalti a un lado u otro, como quién suele lanzar el último tiro en un partido de baloncesto.

En definitiva, este uso de modelos estadísticos avanzadas da un mayor grado de información y da un valor añadido a todas las personas que están disfrutando de esos partidos y aportan información útil sobre los distintos modelos predictivos de estrategias y técnicas utilizadas en baloncesto que ayudarán tanto a profesionales como a aficionados a entender la dinámica del juego.

## 1.2. Objetivos

El principal objetivo del trabajo es poder definir el estilo de juego característico de cada uno de los equipos de la NBA en sus partidos y la influencia en la clasificación.

Este objetivo principal se puede desglosar en los siguientes objetivos específicos:

- 1) Análisis descriptivo de las variables para su posterior análisis, para identificar relaciones entre las variables.
- 2) Identificar estrategias de juego en base a la información disponible sobre el juego de cada equipo en los partidos, tanto locales como visitantes, mediante un análisis de Componentes Principales que nos indique que variables representan las diferentes estrategias.
- 3) Identificar las estrategias seguidas en los partidos siendo el equipo local.
- 4) Identificar las estrategias seguidas en los partidos siendo el equipo visitante.
- 5) Determinar el éxito general de cada estrategia mediante la utilización de análisis clúster.
- 6) Identificar el éxito general en la predicción mediante un análisis discriminante.
- 7) Identificar el peso de las variables en las predicciones mediante el análisis discriminante.

## 1.3. Metodología

En este apartado se va a describir el software utilizado para la realización del proyecto, así como las diferentes técnicas estadísticas que hemos utilizado durante el trabajo.



### 1.3.1. Software utilizado

El software R es el que se ha utilizado para la elaboración del proyecto. R es un ambiente de programación que es gratuito y abierto para cualquier persona, el cuál proporciona una gran variedad de herramientas, dónde puedes desarrollar técnicas estadísticas y gráficos. Para ampliar las funciones básicas de R se pueden instalar librerías, paquetes e incluso definir tus propias funciones para obtener el resultado deseado.

Dentro de R, tenemos una aplicación web llamada RStudio, el cual permite tratar grandes cantidades de datos, estadísticas... Es un entorno de desarrollo integrado (IDE), que se engloba en una aplicación web.

Se han utilizado las siguientes librerías para realizar el proyecto:

- **library (dplyr)**: Esta librería proporciona una variedad de funciones para manipulación de datos, como filtrado, selección, agregación y transformación de datos. En el código, se utiliza la función **select\_if ()** para seleccionar columnas que cumplen una condición específica.
- **library (e1071)**: Esta biblioteca es parte del paquete de R que proporciona herramientas para el análisis de datos basado en el aprendizaje automático y la minería de datos. Esta librería se utiliza para realizar el coeficiente de asimetría y de kurtosis.
- **library (officer)**: Esta librería se utiliza para crear y modificar documentos de Microsoft Office, como Word y PowerPoint, desde R. En el código, se utiliza la función **write.xlsx ()** para escribir los resultados en un archivo de Excel.
- **library (openxlsx)**: Esta biblioteca también se utiliza para leer y escribir archivos de Excel en R, y ofrece más opciones y características que la biblioteca **readxl**. En el código, se utiliza la función **write.xlsx ()** para escribir los resultados en un archivo de Excel.
- **library(corrplot)**: Esta biblioteca proporciona funciones para visualizar matrices de correlación. En el código, se utiliza la función **corrplot()** para crear un gráfico de correlaciones.
- **library(psych)**: Esta biblioteca ofrece diversas funciones para realizar análisis psicométricos y estadísticos. En el código, se utiliza la función **princomp()** para realizar un análisis de componentes principales (PCA) y **principal()** para obtener los componentes principales utilizando el método de Kaiser.
- **library(cluster)**: Esta biblioteca proporciona funciones para realizar análisis de clustering. En el código, se utiliza la función **kmeans()** para realizar el clustering utilizando el algoritmo k-means.

### 1.3.2. Aprendizaje no supervisado

El aprendizaje no supervisado es un tipo de aprendizaje en el cual no tenemos datos etiquetados, es decir, los desconocemos. Para ello, es necesario encontrar los patrones existentes que tenemos en ese conjunto de datos.

Por tomar un ejemplo, en el aprendizaje no supervisado no existe ningún etiquetado ni categorización de los datos. Entonces, vamos a diferenciar el aprendizaje no supervisado con el supervisado, que lo veremos en puntos posteriores.

En el caso de un aprendizaje supervisado, imaginemos que la computadora está intentando aprender distintos deportes, por tanto se le aplicaría etiquetas que dirían que el fútbol tienen 90 minutos de duración, juegan un total de 11 jugadores en cada equipo y no se puede golpear con la mano el balón, mientras que el baloncesto tiene 40 minutos de duración, juegan un total de 5 jugadores por equipo y no se puede tocar con el pie el balón, y por último, el balonmano tiene una duración de 60 minutos, sólo se puede golpear con la mano y se integra por 7 jugadores por equipo.

Una vez tenemos estos datos, la computadora debe identificar con total seguridad basándose en los datos descriptivos de que deporte estamos hablando cuando tenga que identificar que el deporte es el fútbol. Por tanto, si tiene que identificar, por ejemplo, el baloncesto como deporte, podría decir con seguridad que no se juega con los pies, por lo tanto, no es el fútbol, pero como solo tiene una duración de 40 minutos y juegan 5 jugadores por equipo, también puede asegurar que no es el balonmano. Hay una variable respuesta que es conocida.

En cambio, si hablamos de aprendizaje no supervisado, este tipo de etiquetados y categorización no lo tenemos. Por lo tanto, en este caso la computadora no tendrá la idea del concepto del deporte ya que no puede etiquetar las distintas disciplinas, pero se pueden agrupar según su duración, según con que parte del cuerpo se juega o si necesita un complemento como en el tenis, el número de jugadores por equipo... y a partir de ahí, la máquina agrupará todos estos datos según sus diferencias, similitudes, patrones, y se hará un análisis puro de los datos para poder identificar de cual disciplina hablamos.

Dentro del aprendizaje no supervisado tenemos distintas técnicas de aprendizaje, dentro de las cuales utilizaremos para el análisis exploratorio de los datos y el análisis del éxito de las diferentes estrategias el Análisis de componentes principales y el Análisis de Clúster.

#### 1.3.2.1. Análisis de Componentes Principales

El Análisis de Componentes Principales, también conocido como PCA por sus siglas en inglés, es una técnica clásica de aprendizaje no supervisado que tiene múltiples usos. Algunos de ellos son:

- Exploración de datos: Su aplicación permite descubrir las relaciones existentes entre las variables y observaciones analizadas, así como detectar la presencia de datos atípicos.



- Reducción de dimensionalidad: El PCA emplea la correlación entre las variables para generar nuevas variables llamadas componentes principales, las cuales explican la mayor parte de la variabilidad. La cantidad de variabilidad explicada por cada componente principal es superior a la de las subsiguientes, y, además, las componentes principales explican la variabilidad que no es explicada por las componentes anteriores. Estas componentes principales pueden emplearse como variables nuevas, con la ventaja de no presentar correlación entre sí. Al determinar cuántas componentes principales utilizar durante el análisis, es relevante tener en cuenta que las de menor orden contienen la información más importante.

Esta técnica fue inicialmente desarrollada por Pearson y posteriormente por Hotelling, y trata de transformar todo el conjunto de variables originales en otro conjunto de variables no correlacionadas entre sí, es decir, que no tenga repetición ni redundancia en la información. Estas nuevas variables con llamadas conjunto de componentes principales.

Estas nuevas variables son combinaciones lineales de las anteriores y se van formando en función de la importancia en cuanto a la variabilidad total de las muestras.

Explicándolo de forma más técnica, sería lo siguiente:

Consideramos una matriz de datos  $X$  con unas dimensiones de  $N \times P$  donde la  $N$  serían el número de individuos o registros y la  $P$  el número de variables explicativas.

El PCA te permite encontrar un número de componentes principales  $Z$  ( $Z < P$ ) que explican de forma aproximada lo mismo que las  $P$  variables originales.

El primer componente principal contiene la mayor parte de la información del conjunto inicial de datos, la segunda componente principal contiene el máximo de la variación restante, y así sería sucesivamente con el resto de las componentes principales.

El resultado de este análisis de componentes principales sería una gráfica de *scores* y *loadings*. A través del gráfico de *scores* se puede determinar las diferencias y semejanzas entre las muestras y nos sirve para buscar agrupamientos entre las muestras. Mientras que el gráfico de *loadings* te permite determinar la correlación entre las variables.

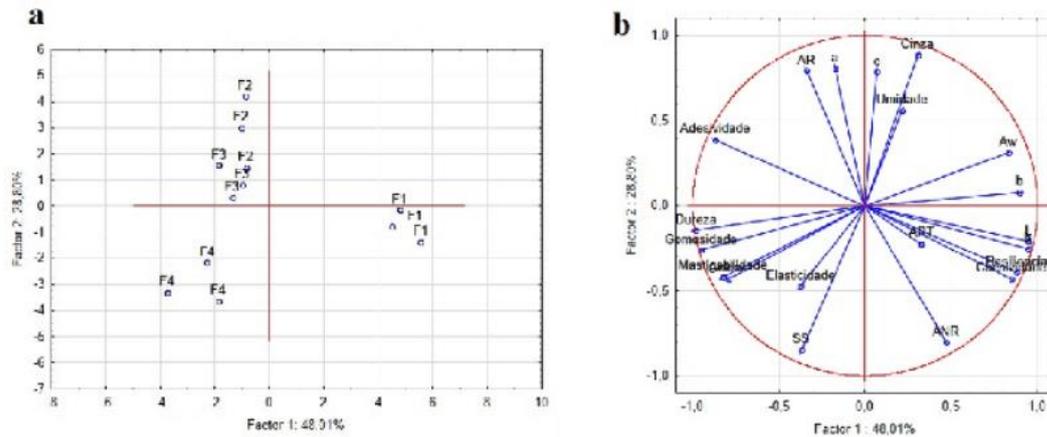


Figura 1. Gráfico de scores (a) y gráfico de loadings (b).

### 1.3.2.2. Análisis Clúster

El análisis clúster es un método estadístico multivariante que consiste en la agrupación de elementos o variables en grupos. El objetivo es que los datos que sacamos de cada grupo sean lo más parecidos posibles entre ellos y lo más diferente posible de los otros grupos. Estos grupos dependen totalmente de los elementos o variables escogidas y la adición o eliminación de variables puede ocasionar grandes diferencias en la solución final.

También es importante tener en cuenta que, en el caso de tener variables o elementos con diferentes unidades de medida, será necesario realizar un paso previo que consiste en la estandarización. Es el método más habitual y consiste en la transformación de los datos, de forma que, tengan unidades de medida similares. En el caso de las variables binarias no se estandarizan y, las variables categóricas, se convertirán en binarias.

Los pasos que seguir en un análisis clúster sería lo siguiente:

Dado un conjunto de variables o elementos  $k = \{k_1, k_2, \dots, k_n\}$  de tamaño  $n$ , se establece un indicador que marcará en qué medida se parecen entre sí cada par de observaciones, bien según la distancia o bien según la similitud. A partir de esto, seguiremos el proceso de **clustering** que consiste en la creación de los grupos en función de la similitud o distancia anteriormente nombrada. Existen distintas técnicas para realizar el análisis clúster, siendo las más utilizadas las que se muestran en la Figura 2.

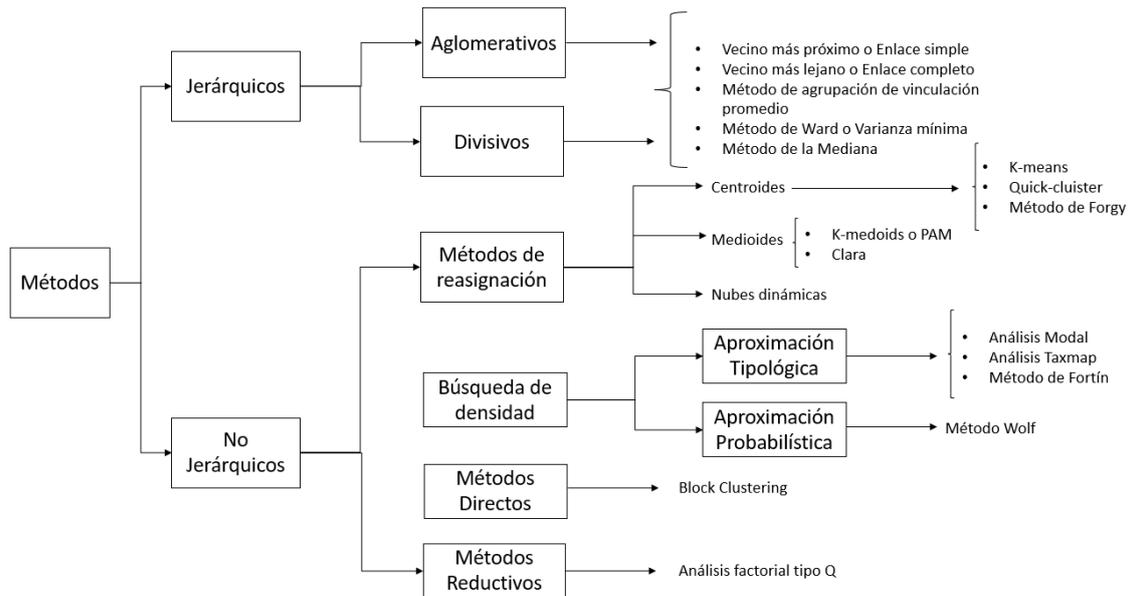


Figura 2: Métodos de Agrupación.

En primer lugar, se diferencia entre métodos jerárquicos y no jerárquicos. Los métodos jerárquicos son aquellos que van generando grupos en cada una de las fases del proceso, hasta que finalmente hace una agrupación óptima. Es decir, en el nivel más alto hay un grupo que se va subdividiendo en más grupos. Aquí, suele ser recomendable que el investigador fije el número de grupos/clústeres que queremos obtener.

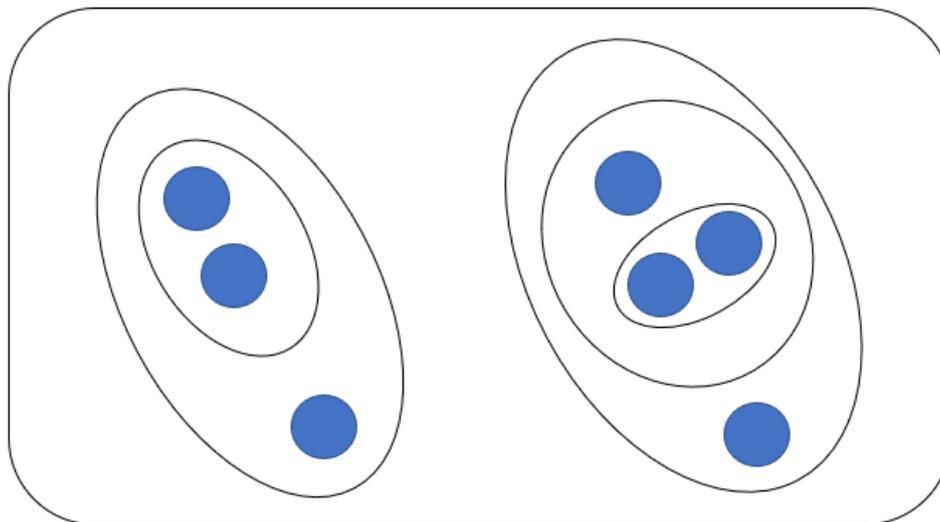
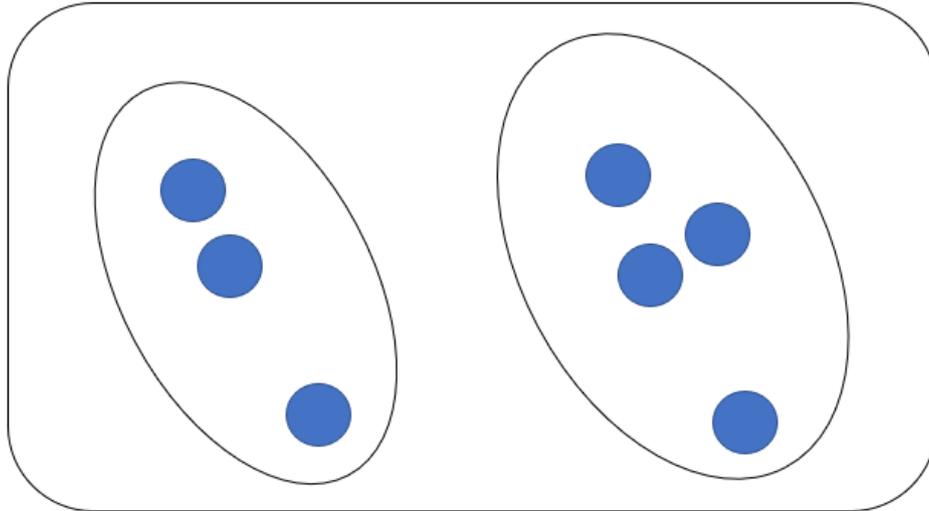


Figura 3: Método no jerárquico.

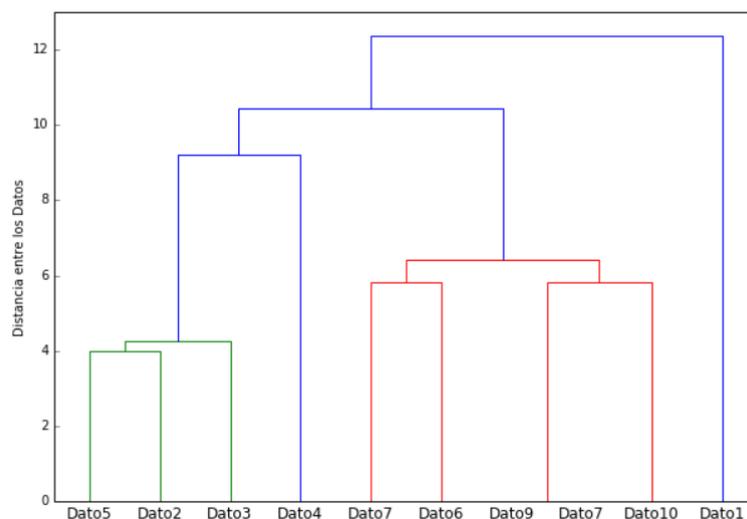
En cuanto a los métodos no jerárquicos, son aquellos que agrupan los elementos o variables en función del número de grupos dados.



**Figura 4:** Método Jerárquico.

Tras haber realizado la creación de los grupos y elegido el método de agrupación para nuestro análisis, la última fase se realiza una descripción de cada uno de los clústers obtenidos y se comparan uno con los otros y, por último, se procede a validar el análisis con los resultados obtenidos.

Específicamente, en este estudio se emplea un enfoque de análisis de clúster jerárquico. Este método genera una estructura jerárquica de divisiones del conjunto de elementos en grupos, representada mediante un árbol de soluciones o dendrograma (ver Figura 5). En el nivel superior del dendrograma se encuentra un único grupo, el cual se subdivide a medida que se desciende en el dendrograma. En este punto, corresponde al analista determinar el número de clústeres deseados.



**Figura 5:** Dendrograma.

En este caso, se ha utilizado el método de Ward para el cálculo de los clústeres. El método de Ward es un algoritmo de análisis de clúster jerárquico que busca minimizar la varianza dentro de cada clúster generado. Comienza considerando cada elemento como un clúster individual y, a medida que avanza, fusiona los clústeres de manera iterativa hasta formar un clúster único que contiene todos los elementos. Durante este proceso, se calcula una medida de disimilitud entre los clústeres, como la suma de cuadrados de las diferencias de los valores de los atributos. El objetivo es fusionar los clústeres que minimicen el aumento en la varianza total.

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

Donde la variable  $x_i$  es el valor de la variable  $X$  para un individuo  $i$ , mientras que la  $n$  sería el número de individuos que tenemos en el clúster.

El método de Ward produce una jerarquía de divisiones de los elementos en clústeres, que se representa mediante un dendrograma. En el dendrograma, cada elemento se encuentra en un nivel inicial, y a medida que se desciende, se forman clústeres más grandes. La altura de cada unión en el dendrograma indica la medida de disimilitud entre los clústeres fusionados. El dendrograma brinda una visualización intuitiva de la estructura jerárquica de los clústeres.

Una vez que se ha determinado el número óptimo de clústeres utilizando el método de Ward, es posible pasar al algoritmo de  $k$ -means. El  $k$ -means es un algoritmo de partición de datos que asigna elementos a clústeres de manera iterativa. Comienza seleccionando  $k$  centroides iniciales y luego asigna cada elemento al clúster representado por el centroide más cercano. Posteriormente, se recalculan los centroides y se repite el proceso hasta que la asignación de clústeres converge.

### 1.3.3. Aprendizaje supervisado

Como hemos explicado anteriormente, el aprendizaje supervisado se basa en el aprendizaje automático que consiste en automatizar mediante distintos algoritmos la identificación de distintos patrones o tendencias en los datos siendo la variable respuesta conocida. Para ello es muy importante tener una gran cantidad de datos que permita identificar estos patrones, además de elegir el algoritmo correcto que permita resolver cada problemática concreta.

Estas técnicas de aprendizaje supervisado suelen utilizar un conjunto de datos de entrenamiento a partir del cual el algoritmo de aprendizaje supervisado crea un modelo predictivo de las variables respuesta a partir de los datos de entrada y un segundo conjunto de datos de validación que serán utilizados para comprobar la eficacia del modelo obtenido,

Dentro del aprendizaje supervisado, tenemos distintos tipos, como el Análisis Discriminante Lineal, Análisis Discriminante de Mínimos Cuadrados Parciales, Random Forest, Redes Neuronales entre otras, pero en este trabajo se utilizará el Discriminante con el objetivo de predecir el éxito o fracaso (ganado o perdido) de los partidos

### 1.3.3.1. Análisis Discriminante

El análisis discriminante consiste en la creación de un modelo predictivo en el que clasifica individuos u objetos según el grupo al que hay más probabilidad que pertenezcan. Esta probabilidad de pertenencia se establece a partir de la observación de diferentes variables. En definitiva, este modelo estadístico busca agrupar en función de ciertas similitudes, de forma que permita cuantificar la probabilidad de pertenecer a un grupo a otro. La diferencia con el análisis de clúster es que los grupos, a priori, son ya conocidos.

El resultado del análisis discriminante lineal dará lugar a una función discriminante combinación lineal de las variables discriminantes que nos permite asignar nuevos individuos en cada uno de los grupos conociendo las diferencias existentes del conjunto de individuos en los distintos grupos y minimice los errores de clasificación.

Para poder hacer un análisis discriminante lineal es importante tener claro una serie de requisitos a seguir:

- 1) En primer lugar, como se ha mencionado antes, hay que tener formados los grupos a analizar, es la diferencia con el análisis de clúster. Es necesario que al menos haya dos grupos, y que en cada uno de éstos contengan al menos dos individuos.
- 2) No debe existir multicolinealidad entre las variables discriminantes a analizar, es decir, no puede haber combinación lineal entre ellas.
- 3) El número máximo de variables discriminantes que se pueden calcular no puede ser mayor al número de individuos menos 2.
- 4) Todas las variables discriminantes tienen que proceder de poblaciones con distribución normal.
- 5) No puede existir bajo ningún concepto varianza igual a cero en ningún grupo, es decir, debe existir dispersión y variabilidad dentro de cada grupo.
- 6) Normalidad Multivariante

Desarrollando más en detalle el análisis discriminante lineal, tenemos un vector de variables discriminantes donde se minimiza la variabilidad dentro de los grupos y se maximiza la variabilidad entre grupos mediante una combinación lineal de estas variables, es decir:

$$D = w_1X_1 + w_2X_2 + \dots + w_pX_p$$

donde  $w_j$  son los coeficientes de ponderación de cada variable clasificadora para  $j = 1, 2, 3, \dots, p$ . Y la  $D$  es nuestra función discriminante. Nuestro objetivo consiste en calcular dichos coeficientes y, para ello, partimos de  $n$  observaciones.

Tras esto queremos clasificar un nuevo elemento en cada uno de los grupos. En este caso, imaginemos que tenemos dos grupos distintos ( $G_1$  y  $G_2$ ). Para clasificar el nuevo elemento

tenemos que calcular su puntuación discriminante ( $d$ ) y, tras esto, será necesario calcular la frontera discriminante:

$$C = \frac{\overline{D}_1 + \overline{D}_2}{2}$$

donde,

$$\overline{D}_1 = w_1\overline{X}_1 + w_2\overline{X}_2 + \dots + w_p\overline{X}_p$$

$$\overline{D}_2 = w_1\overline{X}_1 + w_2\overline{X}_2 + \dots + w_p\overline{X}_p$$

Entonces, tras esto, para clasificar el nuevo elemento, aplicamos el siguiente criterio:

Si  $d < C$  ( $d - C < 0$ ), entonces el nuevo elemento pertenece al grupo 1.

Si  $d > C$  ( $d - C > 0$ ), entonces el nuevo elemento pertenece al grupo 2.

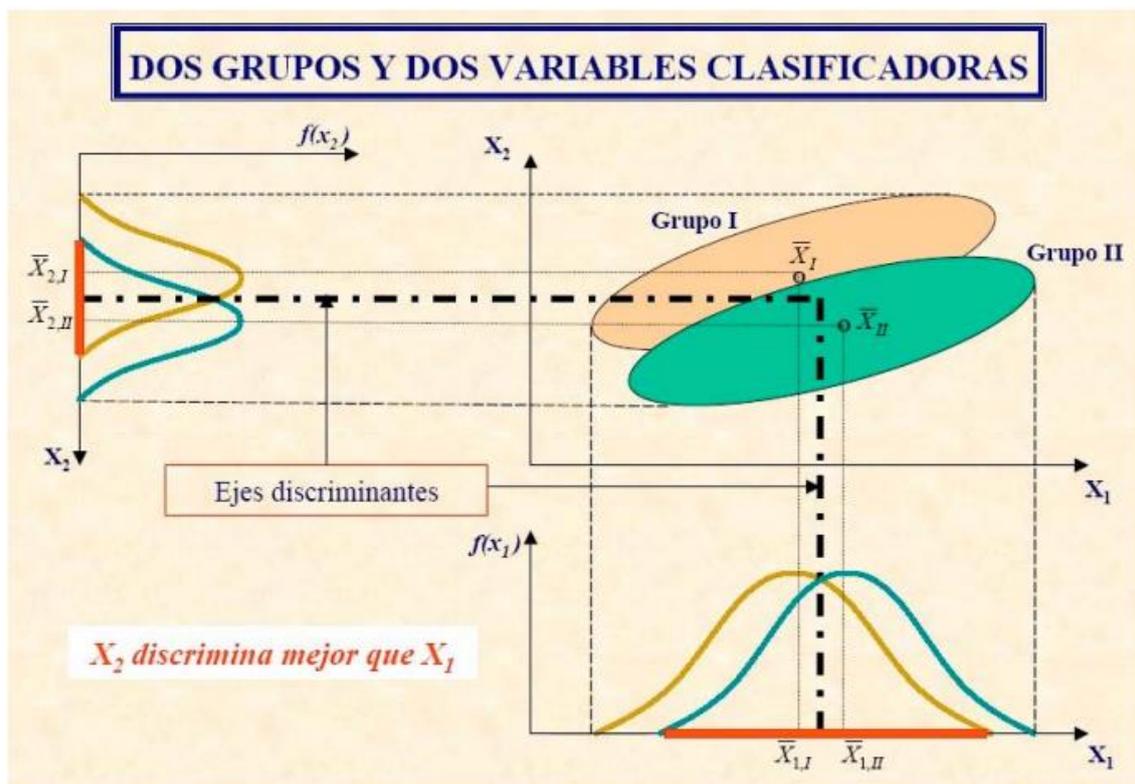


Figura 6: Análisis discriminante lineal.

## Capítulo 2. Análisis de la base de datos

### 2.1. Base de datos

La base de datos utilizada se ha creado en Microsoft Excel y ha recopilado las estadísticas que ha realizado cada uno de los equipos de la NBA en sus partidos como local y visitante en las últimas tres temporadas. Estas estadísticas han sido obtenidas de dos páginas oficiales, ESPN y la web oficial de la NBA.

Una vez dentro de las dos páginas oficiales, tanto de ESPN y la NBA, nos filtramos en cada una de las temporadas, y en cada uno de los equipos, de dónde se sacaron las estadísticas principales de cada partido. Tras sacar cada una de estas estadísticas, se procedió a realizar las estadísticas avanzadas, que consta de 13 variables que serán explicadas en el apartado de la descripción de la base de datos.

La base de datos (BBDD) final constaba de 3638 filas y 84 columnas, donde se recogen tanto las temporadas 19-20, como la 20-21 y la 21-22.

#### 2.1.1. Descripción de la base de datos

La base de datos ha sido creada manualmente y cuenta con las estadísticas convencionales de un partido de baloncesto, y a su vez, las estadísticas avanzadas.

Estas estadísticas avanzadas nos permiten estudiar lo que ocurre en la pista de baloncesto a través del análisis de los datos objetivos y permiten evaluar de manera más precisa la producción del equipo. Para estudiar estas estadísticas contamos con una gran cantidad de variables, tanto cuantitativas como cualitativas que analizaremos en más detalle a continuación.

En total tenemos 3638 registros y un total de 84 variables (8 cualitativas y 78 cuantitativas). La descripción de cada una de ellas se puede ver en el Anexo.

En la primera parte de la Base de Datos (BBDD) se ha recopilado las estadísticas convencionales. La primera parte se compone de los equipos que disputan el partido, seguido de la jornada, que se compone de la fecha del partido. Por lo que las primeras variables que describimos son las siguientes:

- **Equipo (L):** Esta variable indica el equipo que juega de local en la jornada correspondiente. Se trata de una variable **cualitativa nominal**.
- **Equipo (V):** Esta variable indica el equipo que juega de visitante en la jornada correspondiente. Se trata de una variable **cualitativa nominal**.
- **Jornada:** Esta variable indica la fecha en la que se disputa la jornada. Se trata de una variable **cuantitativa discreta**.
- **Mes:** Esta variable indica el mes en el que se disputa la jornada. Se trata de una variable **cuantitativa discreta**.

- **Día:** Esta variable indica el día en el que se disputa la jornada. Se trata de una variable **cuantitativa discreta**.
- **Año:** Esta variable indica el año en el que se disputa la jornada. Se trata de una variable **cuantitativa discreta**.

Equipo (L)
Equipo (V)
Jornada
Mes
Día
Año

*Tabla 1:* Variable incluidas en la base de datos NBA.

En la segunda parte de la BBDD, se han recopilados las variables referentes a los puntos que ha anotado el equipo local y el equipo visitante, después se ha marcado la diferencia que ha habido en el partido, es decir de cuantos puntos ha ganado o ha perdido el equipo local, y en función de este resultado, se ha formulado para que marque si el partido ha sido una victoria o una derrota. La descripción de las variables son las siguientes:

- **PTOS (L):** Esta variable indica los puntos anotados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **PTOS (V):** Esta variable indica los puntos anotados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **Diferencia Marcador (L):** Esta variable indica la diferencia de puntos entre el equipo local y el visitante, es decir, de cuántos puntos ha ganado o ha perdido el equipo local. Se trata de una variable **cuantitativa discreta**.
- **Diferencia Marcador (V):** Esta variable indica la diferencia de puntos entre el equipo local y el visitante, es decir, de cuántos puntos ha ganado o ha perdido el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **V/D (L):** Esta variable indica si el equipo local ha logrado la victoria o ha salido derrotado. Se trata de una variable **cuantitativa binaria**.
- **V/D (V):** Esta variable indica si el equipo visitante ha logrado la victoria o ha salido derrotado. Se trata de una variable **cuantitativa binaria**.

Ptos (L)
Ptos (V)
Diferencia Marcador (L)
Diferencia Marcador (V)
V/D (L)
V/D (V)

*Tabla 2:* Variable incluidas en la base de datos NBA (1).

A continuación, se han recopilado las estadísticas del equipo local en cuanto a tiros de dos puntos y tiros de tres puntos. Y, a su vez los tiros de campo totales, que engloban los dos anteriores. Además, también se ha obtenido las estadísticas de lanzamientos de tiros libres. La descripción de estas variables son las siguientes:

- **T2A (L):** Esta variable indica los tiros de dos puntos anotados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **T2I (L):** Esta variable indica los tiros de dos puntos lanzados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **T2% (L):** Esta variable indica el porcentaje de acierto de tiros de dos puntos por el equipo local. Se calcula dividiendo los TCA por los TCI. Se trata de una variable **cuantitativa continua**.
- **3PA (L):** Esta variable indica los tiros de tres puntos anotados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **3PI (L):** Esta variable indica los tiros de tres puntos lanzados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **3P% (L):** Esta variable indica el porcentaje de acierto de tiros de tres puntos por el equipo local. Se calcula dividiendo los 3PA por los 3PI. Se trata de una variable **cuantitativa continua**.
- **TCA (L):** Esta variable indica los tiros de campo anotados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **TCI (L):** Esta variable indica los tiros de campo lanzados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **TC% (L):** Esta variable indica el porcentaje de acierto de tiros de campo por el equipo local. Se calcula dividiendo los TCA por los TCI. Se trata de una variable **cuantitativa continua**.
- **TLA (L):** Esta variable indica los tiros libres anotados por el equipo local. Se trata de una variable **cuantitativa discreta**.



- **TLI (L):** Esta variable indica los tiros libres lanzados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **TL% (L):** Esta variable indica el porcentaje de tiros libres por el equipo local. Se calcula dividiendo los TLA por los TLI. Se trata de una variable **cuantitativa continua**.
- **T2A (V):** Esta variable indica los tiros de dos puntos anotados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **T2I (V):** Esta variable indica los tiros de dos puntos lanzados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **T2% (V):** Esta variable indica el porcentaje de acierto de tiros de dos puntos por el equipo visitante. Se calcula dividiendo los TCA por los TCI. Se trata de una variable **cuantitativa continua**.
- **3PA (V):** Esta variable indica los tiros de tres puntos anotados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **3PI (V):** Esta variable indica los tiros de tres puntos lanzados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **3P% (V):** Esta variable indica el porcentaje de acierto de tiros de tres puntos por el equipo visitante. Se calcula dividiendo los 3PA por los 3PI. Se trata de una variable **cuantitativa continua**.
- **TCA (V):** Esta variable indica los tiros de campo anotados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **TCI (V):** Esta variable indica los tiros de campo lanzados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **TC% (V):** Esta variable indica el porcentaje de acierto de tiros de campo por el equipo visitante. Se calcula dividiendo los TCA por los TCI. Se trata de una variable **cuantitativa continua**.
- **TLA (V):** Esta variable indica los tiros libres anotados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **TLI (V):** Esta variable indica los tiros libres lanzados por el equipo visitante. Se trata de una variable **cuantitativa discreta**.
- **TL% (V):** Esta variable indica el porcentaje de tiros libres por el equipo visitante. Se calcula dividiendo los TLA por los TLI. Se trata de una variable **cuantitativa continua**.

T2A (L)
T2I (L)
T2% (L)

3PA (L)
3PI (L)
3P% (L)
TCA (L)
TCI (L)
TC% (L)
TLA (L)
TLI (L)
TL% (L)
T2A (V)
T2I (V)
T2% (V)
TCA (V)
TCI (V)
TC% (V)
3PA (V)
3PI (V)
3P% (V)
TLA (V)
TLI (V)
TL% (V)

**Tabla 3:** Variables incluidas en la base de datos NBA (2).

A continuación, se han recopilado las estadísticas más influyentes que se producen durante un partido: los rebotes, diferenciados por defensivos y ofensivos, las asistencias, los robos, los taponos, las pérdidas y las faltas personales. La descripción de estas estadísticas es la siguiente:

- **REBO (L):** Esta variable indica los rebotes ofensivos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.



- **REBD (L):** Esta variable indica los rebotes defensivos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **REBT (L):** Esta variable indica los rebotes totales conseguidos por el equipo local. Se calcula sumando los REBO con los REBD. Se trata de una variable **cuantitativa discreta**.
- **AST (L):** Esta variable indica las asistencias realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **ROB (L):** Esta variable indica los robos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **TAP (L):** Esta variable indica los taponos realizados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **PER (L):** Esta variable indica las pérdidas realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **FP (L):** Esta variable indica las faltas personales realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **REBO (V):** Esta variable indica los rebotes ofensivos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **REBD (V):** Esta variable indica los rebotes defensivos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **REBT (V):** Esta variable indica los rebotes totales conseguidos por el equipo local. Se calcula sumando los REBO con los REBD. Se trata de una variable **cuantitativa discreta**.
- **AST (V):** Esta variable indica las asistencias realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **ROB (V):** Esta variable indica los robos conseguidos por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **TAP (V):** Esta variable indica los taponos realizados por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **PER (V):** Esta variable indica las pérdidas realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.
- **FP (V):** Esta variable indica las faltas personales realizadas por el equipo local. Se trata de una variable **cuantitativa discreta**.

REBO (L)
REBD (L)
REBT (L)
AST (L)



ROB (L)
TAP (L)
PER (L)
FP (L)
REBO (V)
REBD (V)
REBT (V)
AST (V)
ROB (V)
TAP (V)
PER (V)
FP (V)

**Tabla 4:** Variables incluidas en la base de datos NBA (3).

La tercera parte de la BBDD se compone de la estadística avanzada ya que proporciona ventajas para el equipo que la usa, al poder ayudar a obtener datos adecuados en todo tipo de contextos del juego. Esta estadística es la que provoca realmente cambios en la forma de jugar de cada equipo y ya hay muchos parámetros que son indispensables en el estudio del equipo contrario.

En este trabajo nos focalizaremos en cuatro factores de Dean Oliver. Estos factores se componen del porcentaje real de tiro, porcentaje de pérdidas, porcentaje de rebote y frecuencia de los tiros libres. La importancia que le asigna Dean Oliver a cada uno de estos factores es 40%, 25%, 20% y 15%, respectivamente.

La descripción de estas variables sería la siguiente:

- **Porcentaje Efectivo de Tiro % (L):** Este factor mide la eficiencia del equipo local en el tiro de campo, teniendo en cuenta el valor adicional de los triples en comparación con los tiros de dos puntos. Se calcula dividiendo los puntos obtenidos por los tiros de campo efectivos (incluyendo triples) entre el número total de intentos de tiros de campo. Se trata de una variable **cuantitativa continua**.
- **Porcentaje Efectivo de Tiro % (V):** Este factor mide la eficiencia del equipo visitante en el tiro de campo, teniendo en cuenta el valor adicional de los triples en comparación con los tiros de dos puntos. Se calcula dividiendo los puntos obtenidos por los tiros de campo efectivos (incluyendo triples) entre el número total de intentos de tiros de campo. Se trata de una variable **cuantitativa continua**.



- **PER% (L):** Este factor mide la habilidad del equipo local para mantener el control del balón y evitar las pérdidas. Se calcula dividiendo el número de pérdidas de balón del equipo entre el número total de posesiones del equipo. Se trata de una variable **cuantitativa continua**.
- **PER% (V):** Este factor mide la habilidad del equipo visitante para mantener el control del balón y evitar las pérdidas. Se calcula dividiendo el número de pérdidas de balón del equipo entre el número total de posesiones del equipo. Se trata de una variable **cuantitativa continua**.
- **REBO% (L):** Este factor mide la capacidad del equipo local para obtener rebotes ofensivos, lo que le permite tener más oportunidades de anotar. Se calcula dividiendo el número de rebotes ofensivos del equipo entre el número total de rebotes ofensivos disponibles (rebotes ofensivos del equipo más rebotes ofensivos del oponente). Se trata de una variable **cuantitativa continua**.
- **REBO% (V):** Este factor mide la capacidad del equipo visitante para obtener rebotes ofensivos, lo que le permite tener más oportunidades de anotar. Se calcula dividiendo el número de rebotes ofensivos del equipo entre el número total de rebotes ofensivos disponibles (rebotes ofensivos del equipo más rebotes ofensivos del oponente). Se trata de una variable **cuantitativa continua**.
- **TL% (EST AVAN – L):** Este factor mide la frecuencia con la que el equipo local va a la línea de tiros libres y su habilidad para convertir esos tiros. Se calcula dividiendo el número de tiros libres intentados por el equipo entre el número de intentos de tiros de campo. Se trata de una variable **cuantitativa continua**.
- **TL% (EST AVAN – V):** Este factor mide la frecuencia con la que el equipo local va a la línea de tiros libres y su habilidad para convertir esos tiros. Se calcula dividiendo el número de tiros libres intentados por el equipo entre el número de intentos de tiros de campo. Se trata de una variable **cuantitativa continua**.

Porcentaje Efectivo de Tiro % (L)
Porcentaje Efectivo de Tiro % (V)
PER% (L)
PER% (V)
REBO% (L)
REBO% (V)
TL% (EST AVAN – L)

TL% (EST AVAN - V)
--------------------

*Tabla 5:* Variables incluidas en la base de datos NBA (4).

También se han incluido otras variables de estadística avanzada que ayuda a distinguir distintos contextos del juego. Estas estadísticas son las que se muestran en la siguiente tabla. La descripción de las variables es la siguiente:

- **Ritmo (L):** Se refiere a la cantidad de posesiones del equipo local durante un juego. Se trata de una variable **cuantitativa continua**.
- **Ritmo (V):** Se refiere a la cantidad de posesiones del equipo visitante durante un juego. Se trata de una variable **cuantitativa continua**.
- **OER (L):** Mide la eficiencia ofensiva del equipo local. Representa la cantidad de puntos anotados por posesión ofensiva. Se trata de una variable **cuantitativa continua**.
- **OER (V):** Mide la eficiencia ofensiva del equipo local. Representa la cantidad de puntos anotados por posesión ofensiva. Se trata de una variable **cuantitativa continua**.
- **DER (L):** Mide la eficiencia defensiva del equipo local. Representa la cantidad de puntos permitidos por posesión defensiva Se trata de una variable **cuantitativa continua**.
- **DER (V):** Mide la eficiencia defensiva del equipo local. Representa la cantidad de puntos permitidos por posesión defensiva Se trata de una variable **cuantitativa continua**.
- **NET RATING (L):** Es la diferencia entre el OER y el DER del equipo local. Indica la diferencia promedio de puntos a favor o en contra por cada 100 posesiones. Se trata de una variable **cuantitativa continua**.
- **NET RATING (V):** Es la diferencia entre el OER y el DER del equipo visitante. Indica la diferencia promedio de puntos a favor o en contra por cada 100 posesiones. Se trata de una variable **cuantitativa continua**.
- **AST% (L):** Mide el porcentaje de canastas del equipo local que son resultado de una asistencia. Se trata de una variable **cuantitativa continua**.
- **AST% (V):** Mide el porcentaje de canastas del equipo visitante que son resultado de una asistencia. Se trata de una variable **cuantitativa continua**.
- **AST/PER (L):** Mide la relación entre el número de asistencias y el número de pérdidas de balón del equipo local. Se trata de una variable **cuantitativa continua**.
- **AST/PER (V):** Mide la relación entre el número de asistencias y el número de pérdidas de balón del equipo visitante. Se trata de una variable **cuantitativa continua**.
- **REBD% (L):** Mide el porcentaje de rebotes defensivos obtenidos por el equipo local en relación con los rebotes defensivos totales disponibles. Se trata de una variable **cuantitativa continua**.



- **REBD% (V):** Mide el porcentaje de rebotes defensivos obtenidos por el equipo visitante en relación con los rebotes defensivos totales disponibles. Se trata de una variable **cuantitativa continua**.
- **REB% (L):** Mide el porcentaje de rebotes totales obtenidos por el equipo local en relación con los rebotes totales disponibles. Se trata de una variable **cuantitativa continua**.
- **REB% (V):** Mide el porcentaje de rebotes totales obtenidos por el equipo visitante en relación con los rebotes totales disponibles. Se trata de una variable **cuantitativa continua**.
- **Porcentaje de Tiro Real% (L):** Ajusta el porcentaje de tiros de campo de un jugador del equipo local, teniendo en cuenta el valor adicional de los tiros de tres puntos y los tiros libres. Se trata de una variable **cuantitativa continua**.
- **Porcentaje de Tiro Real% (V):** Ajusta el porcentaje de tiros de campo de un jugador del equipo visitante, teniendo en cuenta el valor adicional de los tiros de tres puntos y los tiros libres. Se trata de una variable **cuantitativa continua**.

Ritmo (L)
OER (L)
DER (L)
NET RATING (L)
AST% (L)
AST/PER (L)
REBD% (L)
REB% (L)
Porcentaje de Tiro Real % (L)
Ritmo (V)
OER (V)
DER (V)
NET RATING (V)
AST% (V)
AST/PER (V)

REBD% (V)
REB% (V)
Porcentaje de Tiro Real % (V)

**Tabla 6:** Variables incluidas en la base de datos NBA (5).

Finalmente, la última parte de la BBDD la asistencia de público en los partidos, también a la liga de los partidos, en nuestro caso exclusivamente la NBA y la temporada que era. La descripción de estas variables es la siguiente:

- **Público:** Esta variable indica si en el partido había público. Se trata de una variable **cualitativa binaria**.
- **% Público:** Esta variable indica el porcentaje de público que había en los partidos. Se trata de una variable **cuantitativa continua**.
- **Liga:** Esta variable indica en la liga que pertenece los equipos, en este caso la NBA. Se trata de una variable **cualitativa nominal**.
- **Temporada:** Esta variable indica a qué temporada corresponde los partidos disputados. Se trata de una variable **cualitativa ordinaria**.

Público
% Público
Liga
Temporada

**Tabla 7:** Variables incluidas en la base de datos NBA (6).

A modo de resumen, en la tabla 8 se muestra todas las variables clasificadas según si son cuantitativas o cualitativas:

VARIABLE CUALITATIVA			VARIABLE CUANTITATIVA	
NOMINAL	ORDINARIA	BINARIA	DISCRETA	CONTINUA
<i>Equipo (L), Equipo (V), Liga</i>	<i>Temporada</i>	<i>V/D (L), V/D (V), Público</i>	<i>Jornada, Ptos (L), Ptos (V), Diferencia Marcador (L), Diferencia Marcador (V) T2A (L), T2I (L), 3PA (L), 3PI (L), TCA (L), TCI (L), TLA (L),</i>	<i>T2% (L), 3P% (L), TC% (L), TL% (L), T2% (V), TC% (V), 3P% (V), TL% (V), Porcentaje Efectivo de Tiro</i>

			<p><i>TLI (L), T2A (V), T2I (V), TCA (V), TCI (V), 3PA (V), 3PI (V), TLA (V), TLI (V), REBO (L), REBD (L), REBT (L), AST (L), ROB (L), TAP (L), PER (L), FP (L), REBO (V), REBD (V), REBT (V), AST (V), ROB (V), TAP (V), PER (V), FP (V), Mes, Día, Año</i></p>	<p><i>% (L), Porcentaje Efectivo de Tiro % (V), PER% (L), PER% (V), REBO% (L), REBO% (V), TL% (EST AVAN – L), TL% (EST AVAN - V), Ritmo (L), OER (L), DER (L), NET RATING (L), AST% (L), AST/PER (L), REBD% (L), REB% (L), Porcentaje de Tiro Real % (L), Ritmo (V), OER (V), DER (V), NET RATING (V), AST% (V), AST/PER (V), REBD% (V), REB% (V), Porcentaje de Tiro Real % (V), % Público</i></p>
--	--	--	--	---

**Tabla 8:** Variables Base de Datos.

### 2.1.2. Selección de variables:

Teniendo en cuenta los objetivos marcados de identificar las estrategias de los equipos a partir de los datos, tenemos que seleccionar unas variables en función de una serie de criterios:

- Eliminamos variables, como mes, año, día, jornada ya que no nos sirven como estadísticas en nuestro estudio para determinar las estrategias de nuestros equipos de la NBA.

Por tanto, las variables seleccionadas para hacer los análisis son las siguientes:

Equipo (L)
Equipo (V)
Ptos (L)
Ptos (V)
Diferencia Marcador (L)



Diferencia Marcador (V)
V/D (L)
V/D (V)
TCI (L)
TCI (V)
TCA (L)
TCA (V)
TC% (L)
TC% (V)
T2A (L)
T2I (L)
T2% (L)
3PA (L)
3PI (L)
3P% (L)
TLA (L)
TLI (L)
TL% (L)
T2A (V)
T2I (V)
T2% (V)
3PA (V)
3PI (V)
3P% (V)
TLA (V)
TLI (V)



TL% (V)
REBO (L)
REBD (L)
REBT (L)
AST (L)
ROB (L)
TAP (L)
PER (L)
FP (L)
REBO (V)
REBD (V)
REBT (V)
AST (V)
ROB (V)
TAP (V)
PER (V)
FP (V)
Porcentaje Efectivo de Tiro % (L)
Porcentaje Efectivo de Tiro % (V)
PER% (L)
PER% (V)
REBO% (L)
REBO% (V)
TL% (EST AVAN – L)
TL% (EST AVAN - V)



Ritmo (L)
OER (L)
DER (L)
NET RATING (L)
AST% (L)
AST/PER (L)
REBD% (L)
REB% (L)
Porcentaje de Tiro Real % (L)
Ritmo (V)
OER (V)
DER (V)
NET RATING (V)
AST% (V)
AST/PER (V)
REBD% (V)
REB% (V)
Porcentaje de Tiro Real % (V)
% Público
Temporada

*Tabla 9:* Variables seleccionadas.

### 2.1.2.1. Análisis descriptivo inicial de las estadísticas de los equipos mediante gráficos

Como primer análisis exploratorio de los datos, y antes de empezar a modelar, es importante ver como se distribuyen nuestros datos. Para ello mediante Excel, se van a realizar una serie de tablas dinámicas donde se va a poder ver una serie de gráficos los cuales nos mostrará una primera referencia visual de cómo se comportan las variables seleccionadas. Las gráficas serían las siguientes:

a) Puntos anotados y como los puntos anotados como visitante:

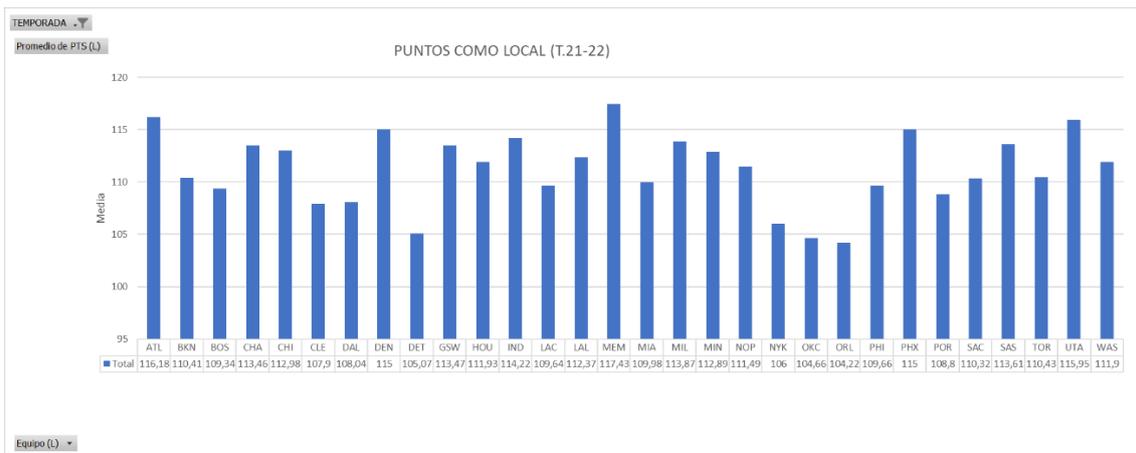


Figura 7: Puntos como local (T.21-22).

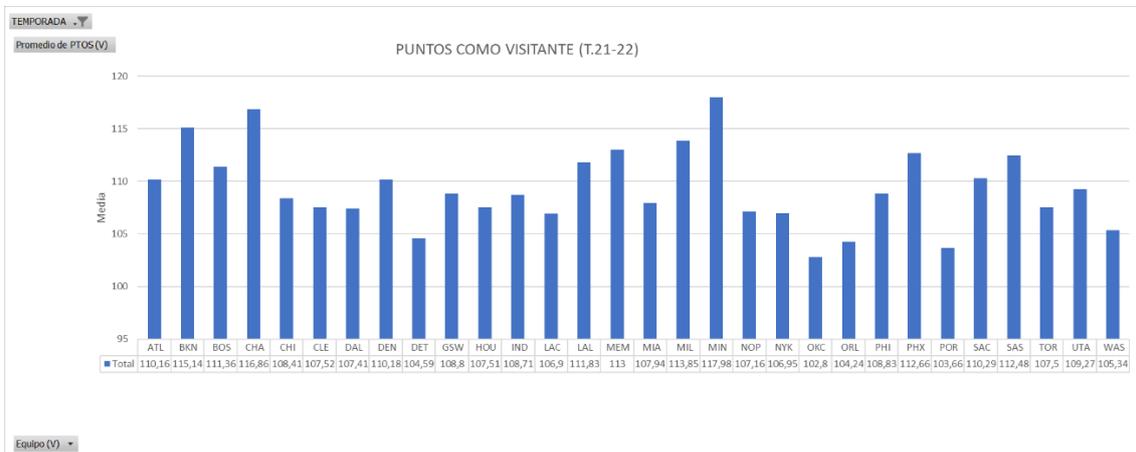


Figura 8: Puntos como visitante (T.21-22).

b) Diferencia en el resultado, tanto como local como visitante:

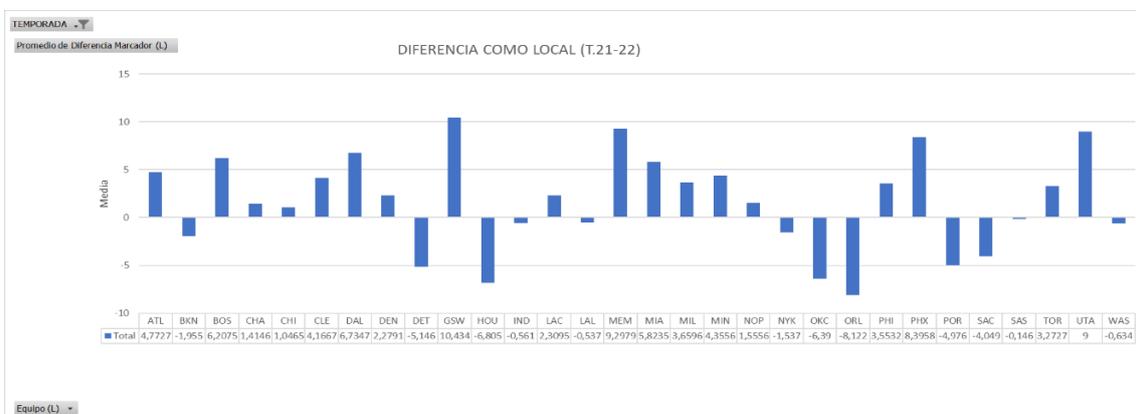


Figura 9: Diferencia como local (T.21-22).



Figura 10: Diferencia como visitante (T.21-22).

Podemos apreciar, por ejemplo, que en equipos como Atlanta, como local tiene una diferencia positiva, mientras que como visitante es negativo, es decir, es más fuerte como local. Por ejemplo, hay equipos como Golden State Warriors Y Phoenix Suns, que tanto como local como visitante, la diferencia es positiva, pero en casa tienen mejores resultados y consiguen victorias más amplias. Y, luego equipos como Brooklyn Nets, que consiguen mejores resultados fuera de casa, que en su propio estadio.

c) T2, T3, TC y TL, tanto como local como visitante:

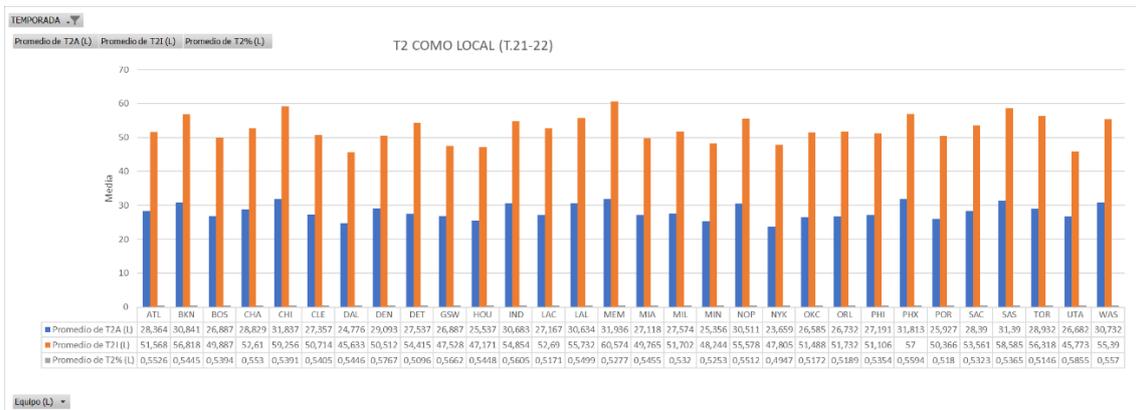


Figura 11: T2 como local (T.21-22).

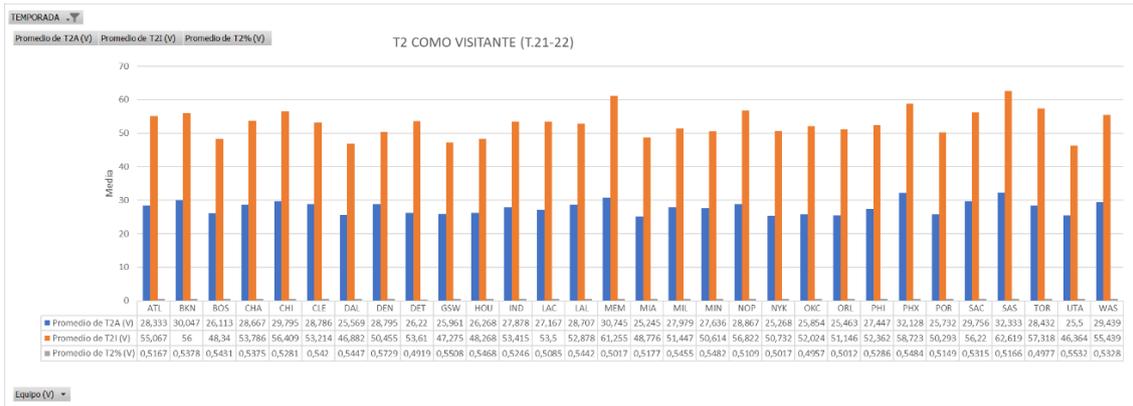


Figura 12: T2 como visitante (T.21-22).

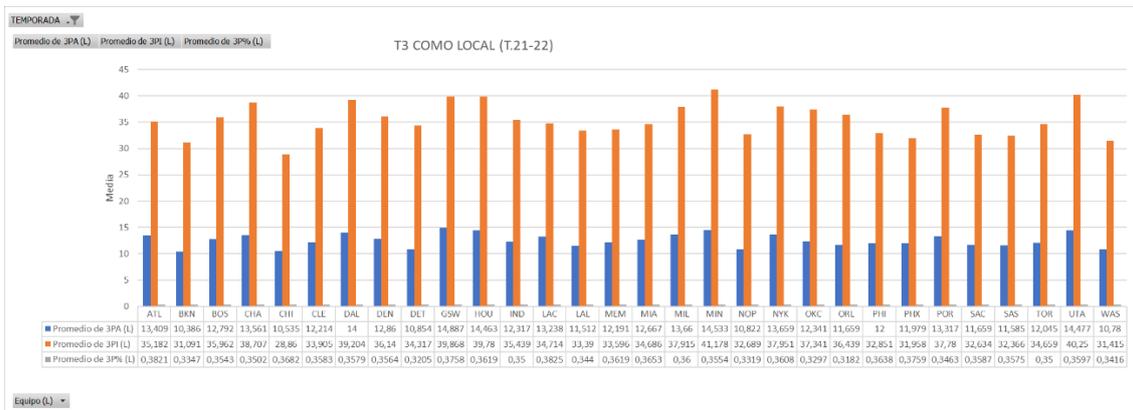


Figura 13: T3 como local (T.21-22).

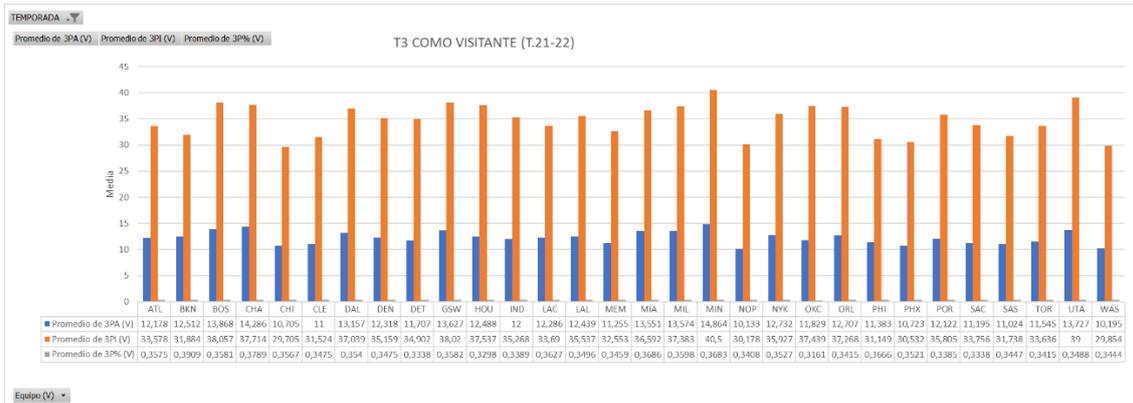


Figura 14: T3 como visitante (T.21-22).

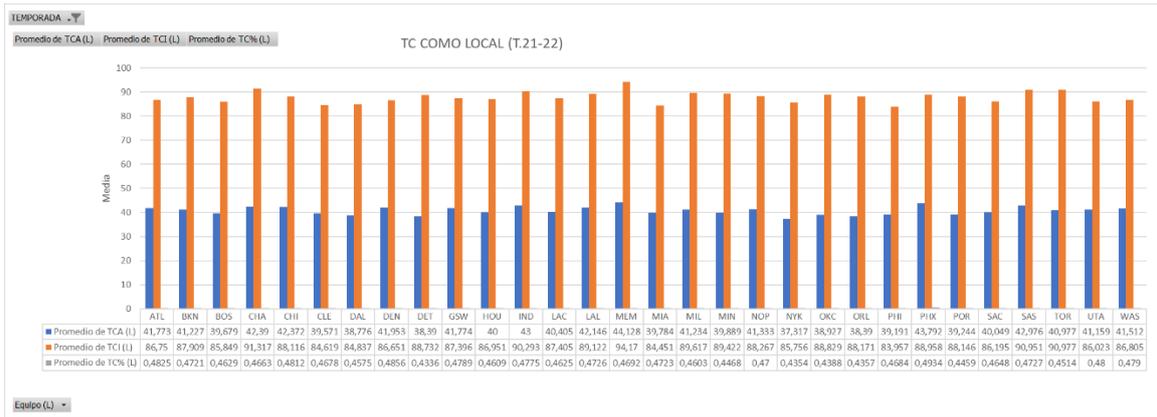


Figura 15: TC como local (T.21-22).

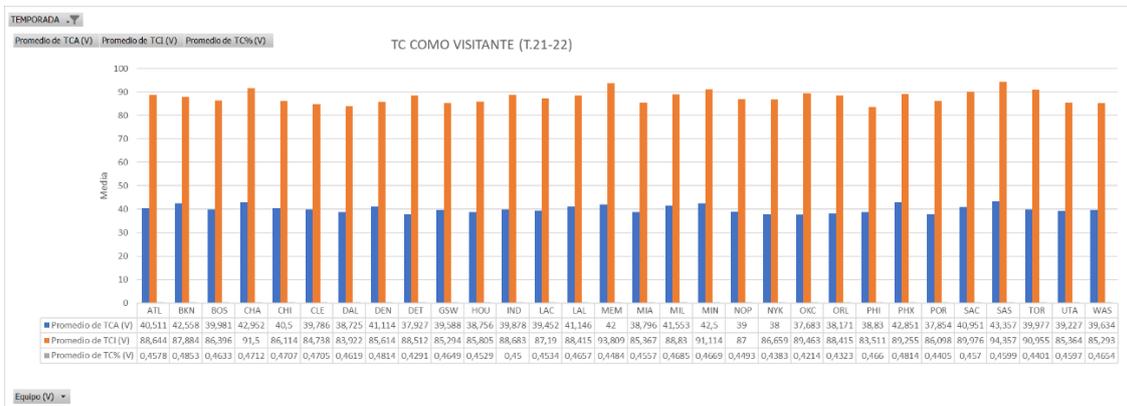


Figura 16: TC como visitante (T.21-22).

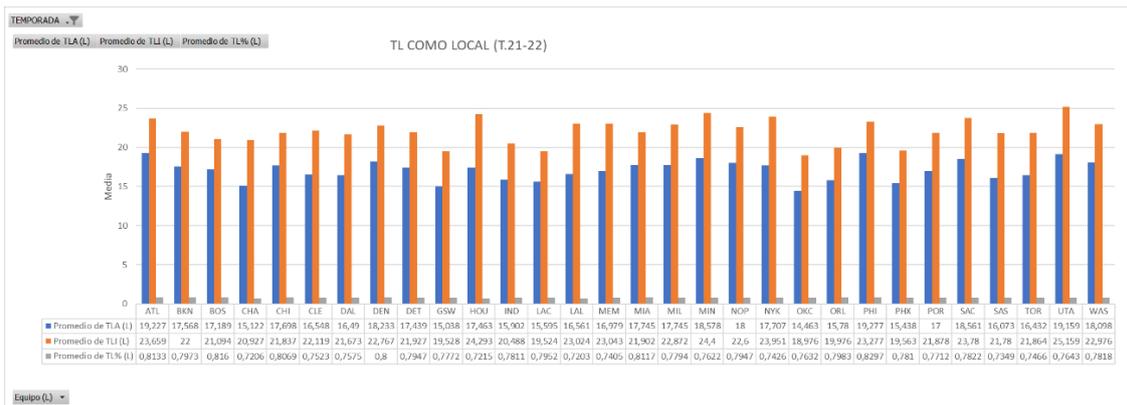


Figura 17: TL como local (T.21-22).

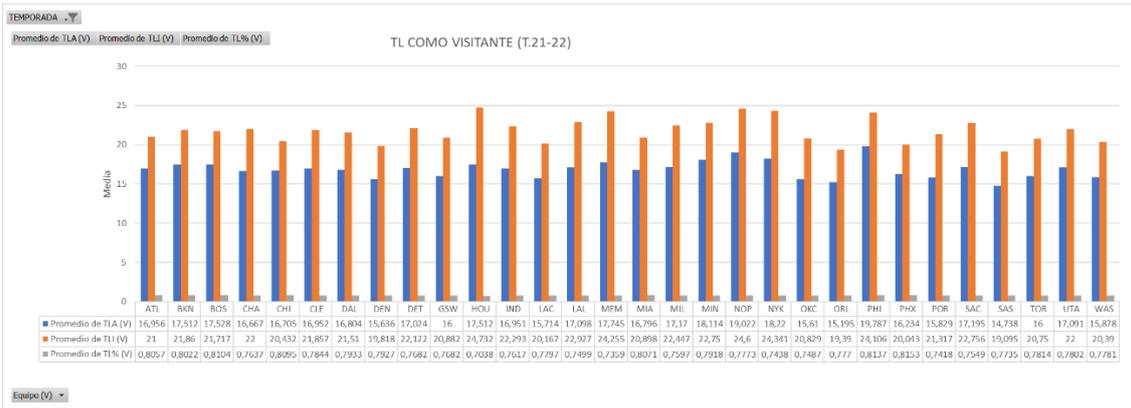


Figura 18: TL como visitante (T.21-22).

Como conclusión se puede apreciar unas características similares en todos los equipos. Esto marca un análisis muy descriptivo de los lanzamientos que se realizan en los partidos, pero se puede ver como los porcentajes en todos los casos son parecidos.

d) Rebotes, clasificados en ofensivos y defensivos, y la suma total de ellos:

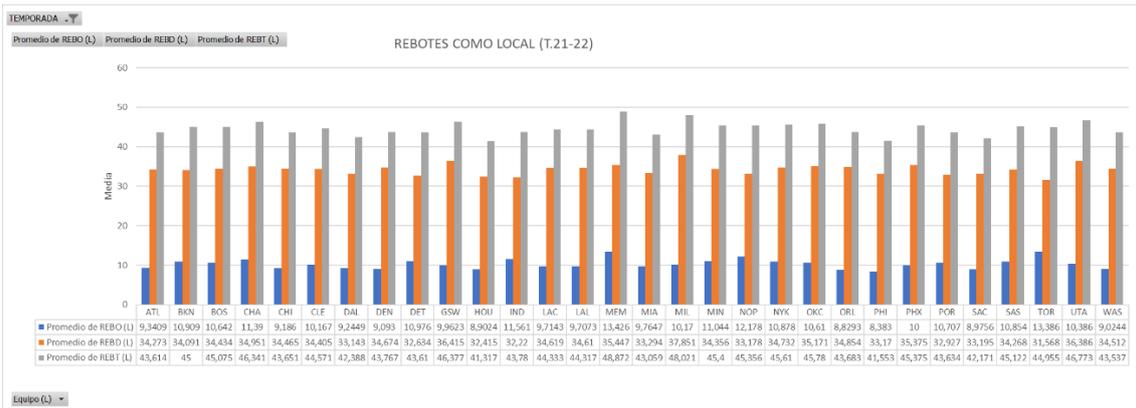


Figura 19: Rebotes como local (T.21-22).

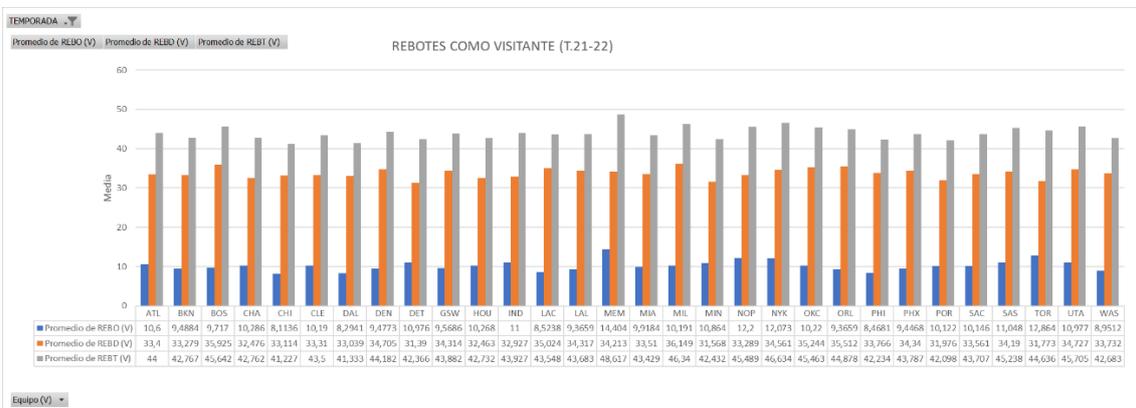


Figura 20: Rebotes como visitante (T.21-22).

En este caso, estamos en una situación similar a la anterior. Todos tienen unos números similares en los rebotes conseguidos. Tanto como local como visitante, los equipos rondan entre los 40-50 rebotes por partido.

Y, por último, mostramos un primer avance de cómo se han desempeñado los equipos en asistencias, robos, tapones, pérdidas y faltas personales, tanto como local como visitante, que como se podrá ver, tienen promedios similares:

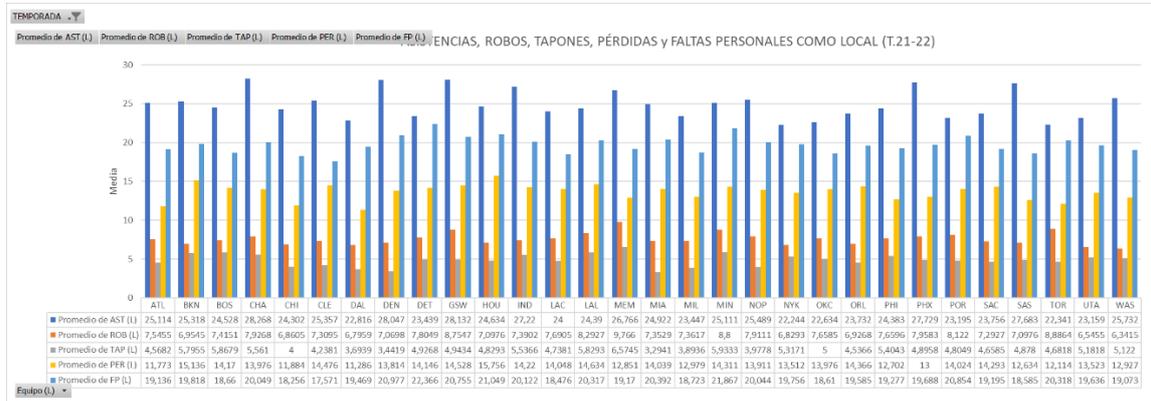


Figura 21: AST, ROB, TAP, PÉR y FP como local (T.21-22).

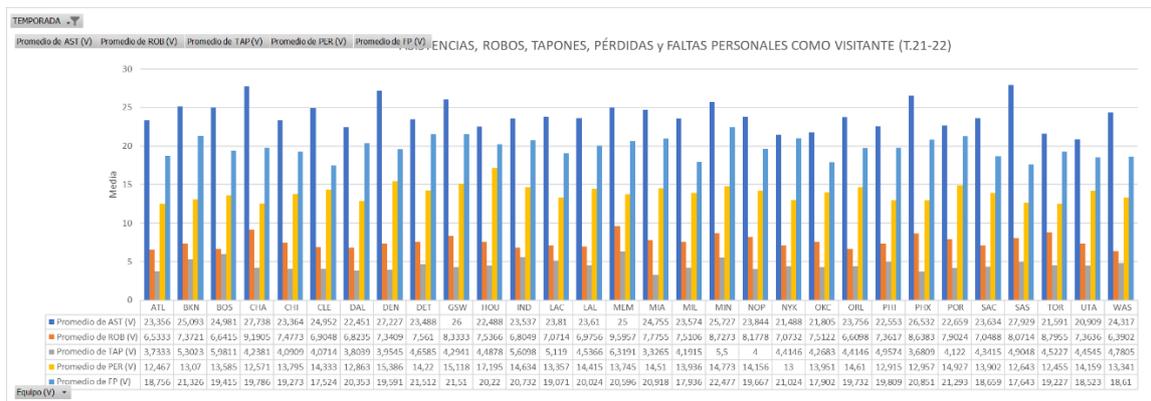


Figura 22: AST, ROB, TAP, PÉR y FP como visitante (T.21-22).

### 2.1.2.2. Análisis descriptivo de las variables. Estudio de la normalidad

En esta parte tratamos de realizar un análisis descriptivo de las variables mediante RStudio. Para eso se realizaron los scripts que se pueden ver en el Anexo. Y se obtuvieron los siguientes resultados:

	Media	Desviación Típica	Mínimo	Máximo	Coefficiente Asimetría	Coefficiente Curtosis
PTS (L)	111,265	12,428	77,000	153,000	0,080	-0,131
PTOS (V)	109,379	12,726	75,000	158,000	0,206	0,122
Diferencia Marcador (L)	1,887	15,320	-48,000	73,000	0,060	0,323
Diferencia Marcador (V)	-1,887	15,320	-73,000	48,000	-0,060	0,323
T2A (L)	28,214	5,391	10,000	46,000	0,063	-0,052
T2I (L)	52,404	8,196	30,000	86,000	0,172	0,123
T2% (L)	0,540	0,075	0,281	0,756	-0,005	-0,003
3PA (L)	12,577	3,660	3,000	27,000	0,330	0,014
3PI (L)	35,443	6,533	18,000	58,000	0,266	-0,107
3P% (L)	0,355	0,082	0,095	0,630	0,104	-0,117
TCA (L)	40,791	5,283	23,000	60,000	0,087	-0,066
TCI (L)	87,847	6,990	67,000	113,000	0,264	0,303
TC% (L)	0,465	0,056	0,288	0,653	0,085	-0,138
TLA (L)	17,107	5,926	3,000	37,000	0,349	-0,052
TLI (L)	22,076	7,219	6,000	49,000	0,365	0,021
TL% (L)	0,775	0,102	0,308	1,000	-0,313	0,253
REBO (L)	10,316	3,804	1,000	25,000	0,467	0,135
REBD (L)	34,281	5,405	16,000	54,000	0,260	-0,001
REBT (L)	44,597	6,664	28,000	70,000	0,315	0,098
T2A (V)	27,852	5,145	12,000	48,000	0,140	0,101
T2I (V)	52,957	8,153	28,000	81,000	0,105	-0,062
T2% (V)	0,528	0,075	0,291	0,796	0,107	0,048
TCA (V)	40,119	5,211	23,000	58,000	0,179	0,128
TCI (V)	87,754	6,918	67,000	120,000	0,288	0,417
TC% (V)	0,458	0,055	0,277	0,687	0,153	0,012
3PA (V)	12,267	3,802	2,000	26,000	0,320	0,033
3PI (V)	34,797	6,564	15,000	57,000	0,045	-0,094
3P% (V)	0,351	0,083	0,077	0,593	0,133	-0,048
TLA (V)	16,875	5,525	4,000	42,000	0,337	0,178
TLI (V)	21,775	6,696	5,000	49,000	0,295	0,078
TL% (V)	0,777	0,104	0,375	1,000	-0,306	0,260
REBO (V)	10,226	3,883	1,000	26,000	0,501	0,270
REBD (V)	33,751	5,314	18,000	54,000	0,035	-0,042
REBT (V)	43,977	6,553	25,000	66,000	0,106	-0,123
AST (V)	24,094	4,998	8,000	41,000	0,225	0,095
ROB (V)	7,580	2,876	0,000	19,000	0,495	0,275
TAP (V)	4,536	2,315	0,000	15,000	0,691	0,573
PER (V)	13,888	3,696	2,000	30,000	0,380	0,469
FP (V)	19,814	4,149	8,000	33,000	0,172	-0,129
AST (L)	24,953	5,052	11,000	41,000	0,288	-0,076
ROB (L)	7,596	2,808	0,000	21,000	0,503	0,613
TAP (L)	4,865	2,478	0,000	16,000	0,489	0,175
PER (L)	13,618	3,790	3,000	25,000	0,264	-0,177
FP (L)	19,725	3,887	8,000	34,000	0,189	0,245
PORCENTAJE EFECTIVO DE TIRO% (L)	53,716	6,625	33,750	75,301	0,156	-0,118
PORCENTAJE EFECTIVO DE TIRO% (V)	52,820	6,709	31,928	79,518	0,244	0,083
PER% (L)	13,469	3,568	3,344	24,752	0,217	-0,147
PER% (V)	13,738	3,537	2,122	27,747	0,323	0,310
REBO% (L)	23,178	7,150	2,564	47,826	0,177	-0,168
REBO% (V)	22,126	6,927	2,222	44,231	0,104	-0,243
TL% (EST AVANZ - L)	19,764	7,500	3,125	48,611	0,546	0,296
TL% (EST AVANZ - V)	19,449	6,829	3,509	50,000	0,415	0,111
RITMO (L)	100,862	5,303	86,480	128,880	0,636	1,639
RITMO (V)	100,997	5,322	86,120	131,760	0,591	1,718
OER (L)	1,104	0,117	0,778	1,521	0,006	-0,104
OER (V)	1,086	0,135	0,710	1,628	0,213	0,016
DER (L)	1,085	0,116	0,708	1,472	0,043	0,069
DER (V)	1,104	0,129	0,747	1,550	0,129	0,043
NET RATING (L)	0,019	0,152	-0,510	0,731	0,064	0,321
NET RATING (V)	-0,018	0,152	-0,723	0,466	-0,059	0,331
AST% (L)	0,611	0,091	0,310	0,868	-0,075	-0,189
AST% (V)	0,600	0,092	0,276	0,878	-0,196	-0,061
AST/PER (L)	2,000	0,794	0,652	7,333	1,875	6,782
AST/PER (V)	1,887	0,766	0,455	11,500	2,638	20,888
REBD% (L)	77,287	7,166	52,727	97,297	-0,171	-0,064
REBD% (V)	76,822	7,150	52,174	97,436	-0,177	-0,168
REB% (L)	50,232	5,103	34,947	67,309	-0,024	-0,261
REB% (V)	49,474	5,032	31,960	65,833	-0,052	-0,076
PORCENTAJE DE TIRO REAL% (L)	57,143	6,236	38,738	76,253	0,127	-0,184
PORCENTAJE DE TIRO REAL% (V)	56,452	7,549	36,909	85,400	0,265	0,018
% PÚBLICO	0,906	0,144	0,000	1,000	-3,241	15,888

Tabla 10: Tabla con el primer análisis descriptivo.

El coeficiente de asimetría mide la asimetría o falta de simetría en la distribución de una variable. Si el coeficiente de asimetría es cero, entonces la distribución es simétrica; si el coeficiente es

positivo, entonces la distribución está sesgada hacia la derecha; y si el coeficiente es negativo, entonces la distribución está sesgada hacia la izquierda.

En mi conjunto de datos, la mayoría de los valores se encuentran en el rango de 0 a 1, lo que indica una asimetría positiva moderada en la distribución de los datos. Algunos valores son cercanos a cero, lo que sugiere una distribución más simétrica en esas variables. Hay valores negativos presentes en la lista, indicando una cola larga hacia la izquierda en esas observaciones.

El coeficiente de curtosis mide la forma de la distribución y su concentración alrededor de la media. Un coeficiente de curtosis igual a cero indica una distribución normal (en forma de campana); un coeficiente de curtosis positivo indica una distribución más concentrada alrededor de la media, con colas más pesadas; y un coeficiente de curtosis negativo indica una distribución menos concentrada alrededor de la media, con colas más ligeras. Además, los valores positivos indican una distribución más apuntada (distribución leptocúrtica) y valores negativos indican una distribución más aplanada (distribución platicúrtica).

En mi conjunto de datos, la mayoría de los valores del coeficiente de curtosis están en el rango de -1 a 1, lo que indica que la mayoría de las distribuciones tienen una forma relativamente normal o cercana a ella, aunque hay algunos valores que indican una distribución más concentrada o dispersa alrededor de la media. Los valores se pueden considerar relativamente normales.

Para poder determinar si los datos siguen una distribución normal, es necesario realizar pruebas de normalidad como la prueba de Shapiro-Wilk o la prueba de Kolmogorov-Smirnov. En este trabajo se ha utilizado un test de normalidad de Kolmogorov-Smirnov en RStudio, en la cual se utiliza la función `ks.test()`.

Es una prueba no paramétrica que determina la bondad de ajuste de dos distribuciones de probabilidad entre sí, es decir, del grado en que la distribución observada difiere de otra distribución teórica normal. La toma de la decisión en el contraste puede llevarse a cabo también mediante el empleo del p-valor asociado al estadístico observado. Si el valor p es mayor que el nivel de significancia (generalmente 0.05), entonces se asume que los datos siguen una distribución normal.

Por ejemplo, si el valor p devuelto es 0.15, entonces podemos concluir que los datos siguen una distribución normal al nivel de significancia del 0.05. En cambio, si el valor p es menor que el nivel de significancia, entonces se rechaza la hipótesis nula de que los datos siguen una distribución normal.

En nuestro caso, la prueba de normalidad se obtuvo los siguientes resultados:

```
Asymptotic one-sample Kolmogorov-Smirnov test

data:  nba_season_21_22_analisis
D = 0.93549, p-value < 2.2e-16
alternative hypothesis: two-sided
```

*Figura 23:* Resultado test Kolmogorov-Smirnov.



El resultado de la prueba de Kolmogorov-Smirnov indica que la distribución de los datos en no sigue una distribución normal, ya que el valor de  $p$  es significativamente menor que 0.05 ( $p\text{-value} < 2.2e-16$ ). Esto sugiere que los datos pueden ser asimétricos y tener colas pesadas, lo que puede afectar la interpretación de las medidas de tendencia central y de dispersión.

Es importante tener en cuenta que la prueba de Kolmogorov-Smirnov no indica necesariamente la distribución exacta de los datos, sino que solo compara la distribución empírica de los datos con una distribución teórica (en este caso, una distribución normal). Por lo tanto, es posible que los datos sigan otra distribución diferente a la normal, pero aun así presenten una buena aproximación a la misma, como podemos apreciar mediante los coeficientes de asimetría y curtosis.

### Capítulo 3. Análisis de resultados

En este apartado, se presentan los resultados obtenidos de manera que podamos dar respuesta a los objetivos planteados al inicio del proyecto, desde las estrategias seguidas en los partidos siendo el equipo local o visitante hasta el estudio de las estrategias mayoritariamente por cada equipo y su influencia en la clasificación.

#### 3.1. Estudio de la estrategia de los equipos de la NBA

Para la consecución de los objetivos 1 y 2 formulados en este trabajo de analizar en función de las estadísticas las estrategias que permiten a los equipos tener más probabilidad de vencer respecto a otro, se comenzó estandarizando las variables, de forma que tengamos una desviación típica de 1 y media 0. Para ello, hemos utilizado la estandarización z-score.

Tras esto, se realizó un análisis de correlación entre las variables que se seleccionaron anteriormente para analizar las relaciones entre ellas, y se obtuvo la siguiente matriz de correlaciones:

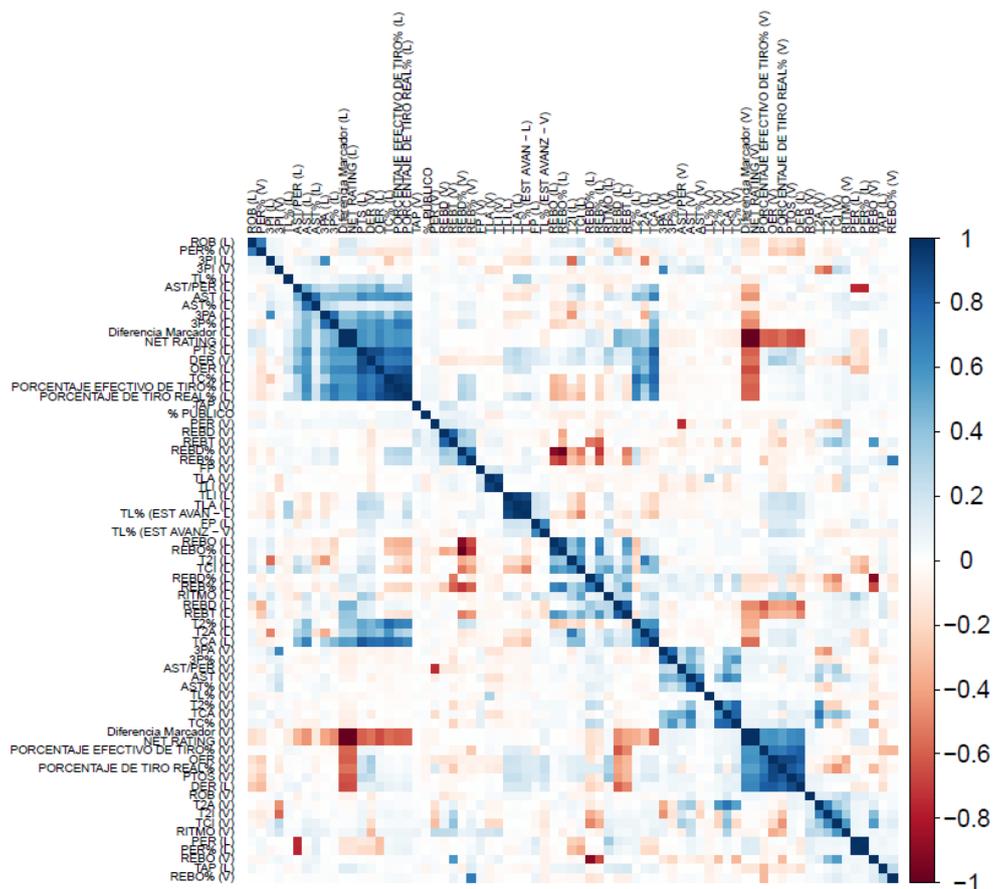


Figura 24: Gráfica correlaciones.

En este caso, podemos sacar varias conclusiones interesantes:

- Todas las variables relacionadas con las estadísticas avanzadas tienen un grado de correlación alto, ya que muchas de ellas están intrínsecas en cada una de ellas. A pesar de que se podría pensar que podría ser redundante alguna de esas estadísticas, no lo es, ya que cada una de ellas nos proporciona un dato distinto.
- Como se puede ver, hay muchas variables que tiene una correlación de 0. Esto sugiere que no hay una relación lineal clara entre ellas.
- Por último, podemos ver algún caso en la que la correlación es negativa, es decir, por ejemplo, en el caso de la variable REBO (V) tiene una correlación negativa próxima o incluso de -1 con la variable REBD% (L). Esto como podemos ver, tiene todo el sentido, ya que, si el equipo rival tiene un mayor número de rebotes ofensivos durante el partido, el porcentaje de rebotes defensivos del equipo local se va a ver disminuido.

Como conclusión, tras el análisis descriptivo de los datos y de la matriz de correlaciones que muestran valores elevados en algunos casos, se decide no descartar ninguna variable, ya que cada una ellas pueden aportar tanto estrategias distintas como puntos de vista distintos dentro del juego.

### 3.1.1. Análisis de componentes principales

Como se ha comentado anteriormente, el análisis de componentes principales es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos multivariados.

Utilizamos la función princomp (), para ello, en primer lugar, nos quedaremos solo con las variables que pertenecen al equipo como local, y posteriormente se analizará como visitante:

#### 3.1.1.1. Equipos como local

Tras seleccionar solo los partidos que han jugado los equipos como local, se obtienen los siguientes resultados en el análisis de PCA:

```
> summary(resultado_pca)
Importance of components:

```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation	2.8955367	2.2643005	1.82778055	1.80158726	1.65258733	1.52187900	1.2796425	1.16293236	1.13711099
Proportion of Variance	0.2330688	0.1425260	0.09286969	0.09022699	0.07591974	0.06438521	0.0455201	0.03759541	0.03594443
Cumulative Proportion	0.2330688	0.3755947	0.46846441	0.55869141	0.63461115	0.69899636	0.7445165	0.78211187	0.81805630
	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17	Comp.18
Standard deviation	1.0753501	1.06182235	1.00318529	0.97854883	0.92717852	0.90176520	0.560374213	0.360414045	0.339482963
Proportion of Variance	0.0321459	0.03134221	0.02797617	0.02661895	0.02389751	0.02260543	0.008729355	0.003611015	0.003203774
Cumulative Proportion	0.8502022	0.88154441	0.90952058	0.93613952	0.96003703	0.98264246	0.991371820	0.994982835	0.998186609
	Comp.19	Comp.20	Comp.21	Comp.22	Comp.23	Comp.24	Comp.25	Comp.26	Comp.27
Standard deviation	0.1536372736	0.117523713	0.0944357959	0.0780383643	0.0682256051	5.810117e-02	4.415278e-02	3.488859e-02	3.383707e-05
Proportion of Variance	0.0006561741	0.000383952	0.0002479129	0.0001692942	0.0001293959	9.384166e-05	5.419284e-05	3.383707e-05	9.999552e-01
Cumulative Proportion	0.9988427827	0.999226735	0.9994746476	0.9996439419	0.9997733378	9.998672e-01	9.999214e-01	9.999552e-01	1
	Comp.28	Comp.29	Comp.30	Comp.31	Comp.32	Comp.33	Comp.34	Comp.35	Comp.36
Standard deviation	2.789765e-02	2.485079e-02	1.467660e-02	3.983763e-08	2.809258e-08	2.058364e-08	1.604489e-08	0	0
Proportion of Variance	2.163521e-05	1.716746e-05	5.987933e-06	4.411771e-17	2.193861e-17	1.177797e-17	7.156475e-18	0	0
Cumulative Proportion	9.999768e-01	9.999940e-01	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1

Figura 25: Resultados PCA.

Tras esto, se utiliza el criterio de Kaiser que es una regla práctica comúnmente utilizada para determinar el número de componentes principales a retener en un análisis de componentes

principales. El criterio establece que solo se deben retener los componentes con valores propios (eigenvalues) mayores que 1.

```
> componentes_kaiser
Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
TRUE FALSE FALSE FALSE
Comp.16 Comp.17 Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25 Comp.26 Comp.27 Comp.28 Comp.29 Comp.30
FALSE FALSE
Comp.31 Comp.32 Comp.33 Comp.34 Comp.35 Comp.36
FALSE FALSE FALSE FALSE FALSE FALSE
```

**Figura 26:** Resultado método Kaiser.

Por lo que tras aplicar el método Kaiser obtuvimos un total de 12 componentes o factores, con autovalores mayores que uno.

<u>Comp.1</u>	<u>Comp.2</u>	<u>Comp.3</u>	<u>Comp.4</u>	<u>Comp.5</u>	<u>Comp.6</u>	<u>Comp.7</u>	<u>Comp.8</u>	<u>Comp.9</u>	<u>Comp.10</u>	<u>Comp.11</u>	<u>Comp.12</u>
2.896	2.264	1.828	1.802	1.653	1.522	1.280	1.163	1.137	1.075	1.062	1.003

**Tabla 11:** Valores método Kaiser.

Las cargas factoriales son un conjunto de coeficientes de cada variable con cada componente que se utilizan en el análisis de componentes principales para describir cómo cada variable está relacionada con los componentes principales.

Estos coeficientes indican la fuerza y dirección de la relación entre cada variable y los componentes principales. Las cargas factoriales se utilizan para interpretar y comprender los componentes principales y para identificar las variables más importantes en cada componente.

En general, una carga factorial alta (cercana a 1 o -1) indica que una variable está altamente relacionada con el componente principal correspondiente, mientras que una carga factorial baja (cercana a 0) indica una relación débil o nula con ese componente.

Estas cargas factoriales también conocidas como loadings, se han obtenido tras aplicar una rotación Varimax que es una técnica utilizada en análisis factorial para mejorar la interpretación de los componentes.

Cuando se realiza el análisis factorial, cada variable original se correlaciona con todas las componentes principales (factores) generadas en el análisis. Sin embargo, cuando hay correlaciones entre las variables originales, puede resultar difícil interpretar las cargas factoriales. La rotación Varimax busca maximizar la varianza de las cargas factoriales de cada variable, haciendo que cada variable se relacione más fuertemente con una sola componente principal.

En resumen, la rotación Varimax ayuda a simplificar la estructura de las cargas factoriales y a facilitar la interpretación de los factores.

Con todo esto, obtenemos la siguiente tabla de cargas factoriales:

	RC1	RC2	RC4	RC6	RC3	RC5	RC7	RC9	RC8	RC10	RC11	RC12
PTS (L)	0,877	0,147	0,188		0,182		0,338					
Diferencia Marcador (L)	0,629			0,706	0,147	-0,680	-0,127				0,174	
T2A (L)	0,488	0,163	-0,169		0,149	-0,799	0,218			-0,410		
T2I (L)		0,367	-0,219		0,259		0,278					
T2% (L)	0,686	-0,184				0,687				-0,665		
3PA (L)	0,494		-0,146			0,907	0,112	0,119		0,473		
3PI (L)		0,162	-0,150			0,141	0,251					
3P% (L)	0,668	-0,132				-0,218				0,679		
TCA (L)	0,841	0,164	-0,274		0,185		0,300					
TCI (L)		0,582	-0,396		0,390	-0,178	0,560					
TC% (L)	0,947	-0,206								-0,116		
TLA (L)			0,971				0,105					0,180
TLI (L)			0,964				0,118					-0,140
TL% (L)			0,169									0,894
REBO (L)	-0,121	0,965										
REBD (L)				0,814			0,226		0,212		-0,347	
REBT (L)		0,581		0,676			0,190		0,156		-0,252	
AST (L)	0,609		-0,161				0,158	0,741				
ROB (L)				0,124		-0,128	0,135				0,870	
TAP (L)		0,106		0,299			0,101		-0,388	0,163		-0,331
PER (L)					-0,967	0,131	0,180					
FP (L)			0,220		-0,118		0,454					
PORCENTAJE EFECTIVO DE TIRO% (L)	0,959	-0,224										
PER% (L)					-0,987							
REBO% (L)		0,963										
TL% (EST AVAN - L)		-0,104	0,972									0,161
RITMO (L)					-0,165		0,928					
OER (L)	0,921	0,109	0,170		0,274		-0,103					0,102
DER (L)	0,102		0,134	-0,912								-0,254
NET RATING (L)	0,631			0,704	0,147	0,169	-0,133					0,172
AST% (L)								0,964				
AST/PER (L)	0,315				0,833			0,330				
REBD% (L)				0,232					0,919			
REB% (L)		0,712	0,156						0,642			
PORCENTAJE DE TIRO REAL% (L)	0,935	-0,236	0,167									
% PÚBLICO				0,110		0,120			-0,102	-0,197	-0,393	0,262

Figura 27: Loadings obtenidos.

Los loadings obtenidos a través de un análisis de componentes principales (PCA) se muestra en la figura o tabla anterior y representan la contribución de cada variable a cada componente principal.

Cada columna representa un componente principal (denotado por RC), y cada fila representa una variable del conjunto de datos original. Los valores en cada celda representan la contribución de cada variable a cada componente principal. Por lo tanto, las variables con valores absolutos más grandes en cada columna son las que tienen una mayor contribución a ese componente principal en particular.

También puede observarse que algunos valores están en blanco (celdas vacías), lo que indica que estas variables no contribuyen significativamente a ese componente principal en particular. En general, los coeficientes de carga se utilizan para interpretar las relaciones entre las variables en un conjunto de datos y pueden ser útiles para seleccionar las variables más importantes para futuros análisis.

Un número positivo indica una relación positiva entre la variable y el componente (es decir, cuando el componente aumenta, la variable también aumenta), mientras que un número negativo indica una relación negativa entre la variable y el componente (es decir, cuando el componente aumenta, la variable disminuye).

Como las 12 componentes son demasiadas, vamos a ver la varianza explicada de cada componente para ver si podemos reducir los componentes principales seleccionados.

Por lo tanto, tras sacar la varianza explicada de los componentes principales, obtenemos la siguiente tabla:

<b>Componentes Principales</b>	<b>Varianza Explicada</b>
<b>RC1</b>	<b>23,30688</b>
<b>RC2</b>	<b>14,2526</b>
<b>RC3</b>	<b>9,286969</b>
<b>RC4</b>	<b>9,022699</b>
<b>RC5</b>	<b>7,591974</b>
<b>RC6</b>	<b>6,438521</b>
<b>RC7</b>	<b>4,55201</b>
<b>RC8</b>	<b>3,759541</b>
<b>RC9</b>	<b>3,594443</b>
<b>RC10</b>	<b>3,21459</b>
<b>RC11</b>	<b>3,134221</b>
<b>RC12</b>	<b>2,797617</b>

*Figura 28:* Varianza explicada de cada componente.

Por lo tanto, tras ver la varianza explicada, nos quedaremos con las 4 primeras componentes que engloban más del 50% de la varianza explicada de cada componente.

Ahora se realiza un estudio de cada componente:

El componente RC1 está compuesto por las siguientes cuatro con los mayores pesos:

1. **PORCENTAJE EFECTIVO DE TIRO% (L) (0,959):** Esta variable indica la eficacia general del equipo al realizar tiros. Un alto valor en RC1 sugiere que el equipo tiene una estrategia ofensiva centrada en lograr un alto porcentaje de tiros efectivos. Esto implica que el equipo busca seleccionar cuidadosamente sus opciones de tiro y enfocarse en la calidad de los lanzamientos para maximizar la eficiencia en la anotación.
2. **TC% (L) (0,947):** Esta variable se refiere al porcentaje de tiros de campo convertidos por el equipo. Un valor alto en RC1 sugiere que el equipo tiene una estrategia de juego enfocada en convertir una alta proporción de sus intentos de tiro. Esto puede indicar un enfoque en la selección inteligente de tiros y en maximizar la precisión de los lanzamientos.
3. **OER (L) (0,921):** Esta variable está relacionada con la eficiencia ofensiva del equipo y su capacidad para generar puntos. Un alto valor en este componente indica que el equipo tiene una estrategia ofensiva eficiente, enfocada en maximizar el rendimiento de sus posesiones ofensivas y aprovechar al máximo las oportunidades de anotación.

4. PTS (L) (0,877): Esta variable representa los puntos anotados por el equipo. Un valor alto en RC1 sugiere que el equipo tiene una estrategia ofensiva agresiva, buscando marcar una alta cantidad de puntos en cada partido.

En base a las variables con los mayores pesos en RC1, se puede concluir que el componente RC1 está relacionado con una estrategia ofensiva centrada en la eficiencia en la anotación y la generación de puntos. El equipo tiende a enfocarse en seleccionar tiros de alta calidad, maximizar el porcentaje de tiros convertidos y ser eficiente en el uso de sus posesiones ofensivas. Esta estrategia implica un enfoque en la toma de decisiones inteligentes en cuanto a los lanzamientos y una buena ejecución en términos de precisión en los tiros.

No se observa una fuerte influencia del juego defensivo en el componente RC1, ya que las variables relacionadas con la defensa (como REBO, PER%, REBO% (RIVAL), ROB, TAP, etc.) tienen pesos más bajos o incluso correlaciones negativas. Sin embargo, esto no significa que el juego defensivo sea irrelevante, sino que, en el contexto de este componente particular, el enfoque principal parece estar en la capacidad ofensiva del equipo.

Como se aprecia en la Figura 29, el componente RC1 refleja una estrategia ofensiva enfocada en la eficiencia en la anotación, la selección inteligente de tiros y la generación de puntos. El equipo busca maximizar su porcentaje de tiros convertidos, tener una alta eficiencia ofensiva y marcar una cantidad significativa de puntos en cada partido.

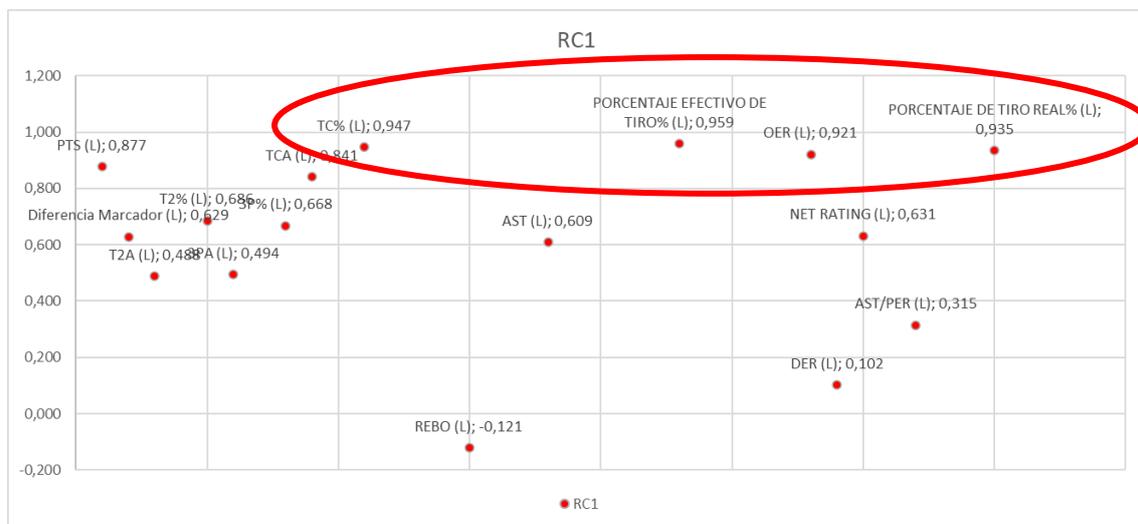


Figura 29: Gráfico componente RC1.

La componente RC2 presenta las siguientes variables con los mayores pesos:

1. REBO (L) (0,965): Esta variable indica el número de rebotes ofensivos capturados por el equipo. Un alto valor en RC2 sugiere que el equipo tiene una estrategia de juego centrada en el rebote ofensivo, buscando obtener segundas oportunidades de anotación y mantener el control del balón en su posesión.
2. REBO% (L) (0,963): Esta variable se refiere al porcentaje de rebotes ofensivos obtenidos por el equipo en relación con los rebotes totales disponibles. Un valor alto en RC2 indica que el equipo tiene una estrategia de juego enfocada en dominar el juego de rebotes

ofensivos, lo que puede ser un indicativo de una fuerte presencia en la pintura y una mayor agresividad en la lucha por los rebotes.

3. TCI (L) (0,582): Esta variable representa el número total de tiros de campo convertidos por el equipo. Un valor alto en RC2 sugiere que el equipo tiene una estrategia ofensiva basada en el volumen de tiros de campo convertidos, buscando generar una alta cantidad de puntos a través de su capacidad para encestar.
4. REB% (L) (0,712): Esta variable indica el porcentaje de rebotes totales obtenidos por el equipo en relación con los rebotes totales disponibles. Un valor alto en RC2 sugiere que el equipo tiene una estrategia de juego enfocada en el dominio del juego de rebotes en general, tanto en la defensa como en el ataque. Esto puede indicar una mayor atención a la lucha por los rebotes y una mayor capacidad para controlar el juego en la pintura.

En términos generales, el componente RC2 refleja una estrategia de juego que se centra en el rebote ofensivo, el volumen de tiros de campo convertidos y el dominio del juego de rebotes en general. El equipo busca obtener segundas oportunidades de anotación a través del rebote ofensivo, generar puntos a través de una alta cantidad de tiros de campo convertidos y controlar el juego en la zona mediante un fuerte juego de rebotes tanto en defensa como en ataque.

No se observa una fuerte influencia de la estrategia de lanzamientos de tres puntos en el componente RC2, ya que las variables relacionadas con los tiros de tres puntos tienen pesos más bajos (o incluso correlaciones negativas). Sin embargo, esto no significa que el equipo no tome tiros de tres puntos, sino que, en el contexto de este componente en particular, el enfoque principal parece estar en el rebote y el volumen de tiros de campo convertidos.

En resumen, como puede apreciarse en la Figura 30 la estrategia asociada con el componente RC2 se enfoca en el rebote ofensivo, el volumen de tiros de campo convertidos y el dominio del juego de rebotes en general. El equipo busca obtener segundas oportunidades de anotación, generar puntos a través de una alta cantidad de tiros convertidos y controlar el juego en la pintura mediante un fuerte juego de rebotes.

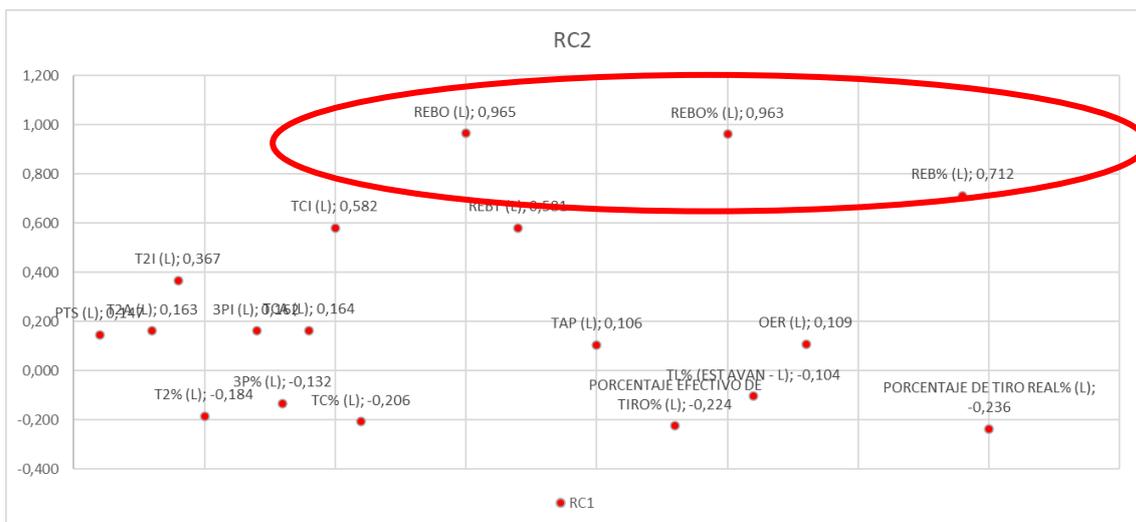


Figura 30: Gráfico componente RC2.

La componente RC3 muestra las siguientes variables con los mayores pesos:

1. PER (L) (-0,967): Esta variable representa el índice de eficiencia del jugador PER (Player Efficiency Rating). Un valor alto en RC3 indica que el equipo tiene una estrategia defensiva enfocada en limitar la eficiencia de los jugadores rivales. Un valor negativo sugiere que el equipo se centra en reducir el rendimiento eficiente de los oponentes, lo que puede implicar una mayor atención a la defensa y a la interrupción del juego ofensivo del equipo contrario.
2. PER% (L) (-0,987): Esta variable se relaciona con el porcentaje de eficiencia del jugador PER. Un valor negativo en RC3 indica que el equipo busca disminuir la eficiencia de los jugadores rivales, lo que puede ser una estrategia defensiva clave.
3. AST/PER (L) (0,833): Esta variable se refiere a la relación entre las asistencias y el índice de eficiencia del jugador PER. Un alto valor en RC3 sugiere que el equipo tiene una estrategia defensiva enfocada en limitar la capacidad de los jugadores rivales para crear jugadas de asistencia y, por lo tanto, disminuir su eficiencia general.
4. OER (L) (0,274): Esta variable representa la eficiencia ofensiva del equipo. Un valor alto en RC3 indica que el equipo tiene una estrategia defensiva centrada en limitar la eficiencia ofensiva del equipo contrario, buscando reducir su capacidad de anotación.

En general, la componente RC3 muestra una estrategia de juego defensiva, donde el equipo se enfoca en limitar la eficiencia y el rendimiento ofensivo de los jugadores rivales. Esto se refleja en los pesos negativos de las variables PER (L), PER% (L) y en el alto peso positivo de AST/PER (L). Además, el equipo busca reducir la eficiencia ofensiva del equipo contrario, como se indica en el peso positivo de OER (L).

No se observa una fuerte influencia de las variables relacionadas con los tiros de campo en esta componente, lo que sugiere que la estrategia principal está centrada en el juego defensivo y en la limitación del rendimiento de los jugadores rivales.

En resumen, como puede apreciarse en la Figura 31, la estrategia asociada con la componente RC3 se enfoca en la defensa y en la limitación de la eficiencia y el rendimiento ofensivo de los jugadores rivales. El equipo busca disminuir la eficiencia general de los jugadores contrarios y reducir la eficiencia ofensiva del equipo contrario en general. Esta componente indica una estrategia defensiva clave para el éxito del equipo.



**Figura 31:** Gráfico componente RC3.

La componente RC4 muestra las siguientes variables con los mayores pesos:

1. TLA (L) (0,971): Esta variable se refiere a los tiros libres intentados por el equipo. Un alto valor en RC4 indica que el equipo tiene una estrategia ofensiva centrada en obtener y aprovechar los tiros libres, lo que sugiere una preferencia por atraer faltas y convertir desde la línea de tiros libres para anotar puntos.
2. TLI (L) (0,964): Esta variable se relaciona con los tiros libres convertidos por el equipo. Un alto valor en RC4 sugiere que el equipo tiene una estrategia ofensiva eficiente en la conversión de tiros libres, lo que indica una habilidad para capitalizar las oportunidades desde la línea de tiros libres.
3. FP (L) (0,220): Esta variable representa las faltas personales cometidas por el equipo. Un valor positivo en RC4 indica que el equipo puede tener una estrategia agresiva en el juego, buscando generar contacto y atraer faltas del equipo contrario para obtener tiros libres y aprovecharlos.
4. TL% (EST AVAN – L) (0,972): Esta variable se refiere al porcentaje de tiros libres convertidos en situaciones avanzadas. Un alto valor en RC4 sugiere que el equipo tiene una estrategia ofensiva efectiva en situaciones de presión, donde logra mantener una alta eficiencia en la conversión de tiros libres.

En general, la componente RC4 muestra una estrategia de juego ofensiva basada en la obtención y conversión de tiros libres. El equipo busca atraer faltas y capitalizar desde la línea de tiros libres para anotar puntos de manera eficiente. La estrategia se refleja en los pesos altos y positivos de las variables TLA (L), TLI (L), FP (L) y TL% (EST AVAN – L).

No se observa una fuerte influencia de las variables relacionadas con los tiros de campo en esta componente, lo que sugiere que la estrategia principal está centrada en el juego ofensivo alrededor de los tiros libres.

En resumen, la estrategia asociada con la componente RC4 se enfoca en una estrategia ofensiva eficiente en el aprovechamiento de los tiros libres, como puede apreciarse en la Figura 31. El equipo busca obtener y capitalizar los tiros libres, aprovechando las oportunidades desde la línea de tiros libres para anotar puntos de manera efectiva. Esta componente indica una estrategia ofensiva clave basada en la habilidad para convertir tiros libres.

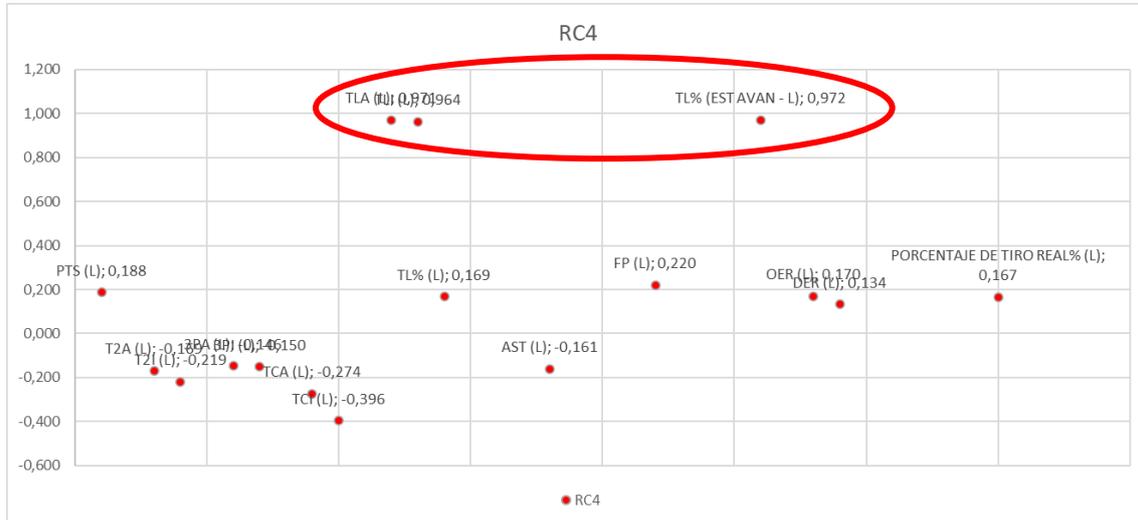


Figura 32: Gráfico componente RC4.

Por tanto, tras ver los distintos componentes uno a uno, obtendríamos las siguientes estrategias:

	<b>Variables</b>	<b>Estrategia</b>
<b>Comp.1</b>	Porcentaje efectivo de tiro % (L), TC% (L), OER (L), PTS (L)	Eficacia en la anotación y la eficiencia ofensiva.
<b>Comp.2</b>	REBO (L), REBO% (L), REB% (L), TCI (L)	Rebote ofensivo, el volumen de tiros de campo convertidos y el dominio del juego de rebotes en general. Búsqueda de segundas oportunidades de anotación.
<b>Comp.3</b>	PER (L), PER% (L), AST/PER (L), OER (L)	Enfoque en la defensa y la limitación de la eficiencia y el rendimiento ofensivo de los jugadores rivales.
<b>Comp.4</b>	TLA (L), TLI (L), FP (L), TL% (EST AVAN - L)	Estrategia ofensiva eficiente en el aprovechamiento de los tiros libres

Tabla 12: Estrategias seguidas cada componente.

### 3.1.1.2. Equipo como visitante

A continuación, se procede a seleccionar sólo los partidos que han jugado los equipos como visitante, y se realizó de nuevo un PCA. Los resultados iniciales sin aplicar el criterio a autovalores mayores que 1 se muestran en la Figura 33.

> summary(resultado\_pca)

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	2.2473224	2.0829685	1.88201869	1.67956926	1.55887401	1.47033954	1.33030443
Proportion of Variance	0.1403966	0.1206122	0.09846316	0.07841908	0.06755351	0.06009816	0.04919579
Cumulative Proportion	0.1403966	0.2610088	0.35947199	0.43789107	0.50544458	0.56554273	0.61473853
	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14
Standard deviation	1.27973206	1.22026557	1.10707048	1.06403448	1.05639599	1.00530839	0.99682809
Proportion of Variance	0.04552647	0.04139373	0.03407034	0.03147294	0.03102268	0.02809471	0.02762272
Cumulative Proportion	0.66026500	0.70165873	0.73572907	0.76720201	0.79822469	0.82631940	0.85394212
	Comp.15	Comp.16	Comp.17	Comp.18	Comp.19	Comp.20	Comp.21
Standard deviation	0.9806136	0.96275815	0.93895015	0.92592006	0.84621089	0.78211852	0.36511511
Proportion of Variance	0.0267314	0.02576679	0.02450817	0.02383268	0.01990596	0.01700478	0.00370583
Cumulative Proportion	0.8806735	0.90644031	0.93094848	0.95478116	0.97468712	0.99169190	0.99539773
	Comp.22	Comp.23	Comp.24	Comp.25	Comp.26	Comp.27	Comp.28
Standard deviation	0.343152868	0.1357117458	0.0879752258	0.0863923006	0.0762464324	5.755400e-02	0.003273416
Proportion of Variance	0.003273416	0.0005119892	0.0002151526	0.0002074799	0.0001616088	9.208245e-05	0.0000448166
Cumulative Proportion	0.998671146	0.9991831351	0.9993982877	0.9996057676	0.9997673764	9.998595e-01	0.9999701833
	Comp.29	Comp.30	Comp.31	Comp.32	Comp.33	Comp.34	Comp.35
Standard deviation	4.869179e-02	0.0401519385	3.275041e-02	3.941737e-08	3.856568e-08	3.699704e-08	0.0000448166
Proportion of Variance	6.590789e-05	0.0000448166	2.981668e-05	4.319180e-17	4.134547e-17	3.805045e-17	0.000000e+00
Cumulative Proportion	9.999254e-01	0.9999701833	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
	Comp.34	Comp.35	Comp.36				
Standard deviation	2.364013e-08	2.243368e-08	2.166377e-08				
Proportion of Variance	1.553551e-17	1.399030e-17	1.304649e-17				
Cumulative Proportion	1.000000e+00	1.000000e+00	1.000000e+00				

Figura 33: Resultados PCA.

Posteriormente se aplicó el criterio de Kaiser para retener los componentes con valores propios (eigenvalues) mayores que 1, obteniéndose un total de 13 componentes como se muestra en la Figura 34, con sus correspondientes cargas factoriales (Tabla 12)

> componentes\_kaiser

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17	Comp.18	Comp.19
TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									

Figura 34: Resultado método Kaiser.

Tras aplicar el método Kaiser obtuvimos tenemos un total de 13 componentes o factores, con autovalores mayores que uno con

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13
2.247	2.083	1.882	1.680	1.559	1.470	1.330	1.280	1.220	1.107	1.064	1.056	1.005

Tabla 13: Cargas Factoriales de las Componentes.

	RC1	RC2	RC3	RC9	RC5	RC4	RC8	RC12	RC7	RC6	RC10	RC11	RC13
PTOS (V)	0,966		0,163										
Diferencia Marcador (V)	0,648						0,746						
T2A (V)		0,758	0,353			-0,493		-0,176	0,135				
T2I (V)			0,628			-0,718			0,168				
T2% (V)		0,920	-0,228			0,146		-0,146					
TCA (V)		0,801	0,369					0,391	0,185		0,101		
TCI (V)	-0,125		0,903	0,110	-0,166				0,270				
TC% (V)		0,854	-0,193					0,426					
3PA (V)						0,587		0,774			0,168		
3PI (V)			0,172		-0,108	0,938					0,112		
3P% (V)		0,141						0,958			0,128		
TLA (V)					0,943							0,232	
TLI (V)					0,946								
TL% (V)					0,118							0,791	
REBO (V)		-0,248	0,638	0,238					0,111		-0,177	-0,171	-0,164
REBD (V)				0,866									
REBT (V)		-0,110	0,397	0,843									
AST (V)		0,480	0,166					0,301	0,154		0,766		
ROB (V)		0,134	0,429	-0,359		0,148		-0,103	-0,201			0,294	
TAP (V)				0,344		0,113						0,306	-0,354
PER (V)				0,120					-0,938				
FP (V)		0,160		-0,141	0,411			0,135	-0,133			-0,421	
PORCENTAJE EFECTIVO DE TIRO% (V)	0,801						0,160			-0,328			
PER% (V)	-0,254	0,129	-0,115			0,166			-0,112			-0,204	-0,503
REBO% (V)			0,138	-0,110						0,861			-0,204
TL% (EST AVANZ - V)	0,248						-0,155						-0,139
RITMO (V)	-0,155	0,159	0,656		0,280		0,158		-0,371		0,183		0,156
OER (V)	0,967		-0,120						0,146				
DER (V)	0,251		-0,118				-0,939						
NET RATING (V)	0,647						0,747						
AST% (V)						0,128					0,956		
AST/PER (V)		0,254	0,129					0,127	0,816		0,354		
REBD% (V)			-0,131	0,501						0,472			0,413
REB% (V)				0,281						0,928			0,153
PORCENTAJE DE TIRO REAL% (V)	0,909		-0,328						-0,163				
% PÚBLICO				0,108		0,128					-0,105		0,614

Figura 35: Loadings obtenidos.

De las 13 componentes sólo con las 5 primeras obteníamos un porcentaje de varianza explicada del 50 %. Por lo tanto, nos quedamos con las 5 primeras, ya que muestran el 50% de la variable explicada.

Componentes Principales	Varianza Explicada
RC1	14,03966
RC2	12,06122
RC3	9,846316
RC4	7,841908
RC5	6,755351
RC6	6,009816
RC7	4,919579
RC8	4,552647
RC9	4,139373
RC10	3,407034
RC11	3,147294
RC12	3,102268
RC13	2,809471

Figura 36: Varianza explicada de cada componente.



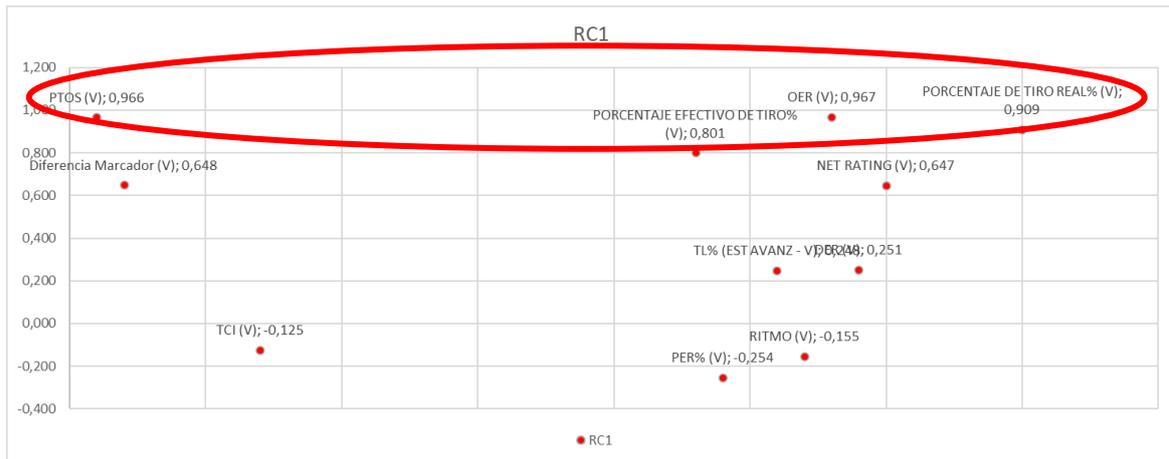
A continuación, procedemos al análisis de las variables que componen cada componente y que contribuirán de definir las posibles estrategias de los equipos visitantes.

La componente RC1 presenta las siguientes variables con los mayores pesos:

1. PTS (V) (0,966): Esta variable representa la puntuación del equipo. Un valor alto en la componente 1 (RC1) indica una fuerte correlación entre la puntuación del equipo y el éxito general. Esto sugiere que la estrategia ofensiva del equipo se centra en anotar muchos puntos para lograr la victoria.
2. PORCENTAJE DE TIRO REAL% (V) (0,909): Esta variable se refiere al porcentaje de tiro real del equipo visitante. Un valor alto en RC1 sugiere que el equipo visitante tiene un buen rendimiento en términos de eficiencia de tiro, teniendo en cuenta los diferentes tipos de lanzamientos realizados.
3. PORCENTAJE EFECTIVO DE TIRO% (V) (0,801): Esta variable representa el porcentaje efectivo de tiros realizados por el equipo. Un valor alto en RC1 sugiere que el equipo se enfoca en una eficiencia en la anotación, seleccionando tiros de alta calidad y maximizando la efectividad en los lanzamientos.
4. OER (V) (0,967): Esta variable se refiere a la eficiencia ofensiva del equipo. Un valor alto en RC1 indica que el equipo tiene una estrategia ofensiva efectiva para generar puntos y mantener una alta eficiencia en sus posesiones ofensivas.

Además, la componente RC1 muestra correlaciones positivas con NET RATING (V) (0,647). Esto sugiere que el equipo también presta atención a mantener una diferencia favorable en el marcador y un buen índice de eficiencia en ambos extremos de la cancha.

En resumen, la estrategia asociada con la componente RC1 implica que el equipo visitante tiene una alta capacidad para anotar puntos y tiende a tener una ventaja en el marcador durante sus partidos. El equipo visitante también muestra una alta eficiencia en sus tiros, tanto en términos de porcentaje efectivo de tiro como de porcentaje de tiro real. Además, el equipo visitante demuestra un buen desempeño tanto en la eficiencia ofensiva como en la eficiencia defensiva. En la Figura 37 se muestra las variables relacionadas con el componente 1.



**Figura 37:** Gráfico componente RC1.

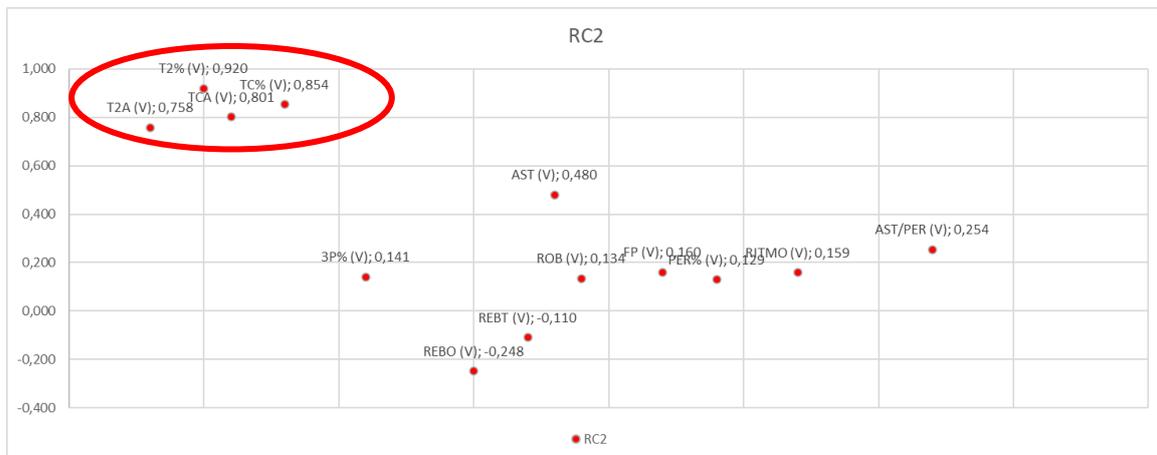
Las variables que más cargaban en la segunda componente (RC2) son las siguientes:

1. T2A (V) (0,758): Esta variable se refiere a los intentos de lanzamiento de dos puntos del equipo visitante. Un valor alto en RC2 sugiere que el equipo visitante tiene una propensión a intentar muchos lanzamientos de dos puntos en sus partidos.
2. T2% (V) (0,920): Esta variable representa el porcentaje de acierto en los lanzamientos de dos puntos del equipo visitante. Un valor alto en RC2 indica que el equipo visitante tiene una alta eficacia en sus tiros de dos puntos.
3. TCA (V) (0,801): Esta variable se refiere a los intentos de campo totales del equipo visitante. Un valor alto en RC2 sugiere que el equipo visitante tiene una tendencia a realizar muchos intentos de campo durante sus partidos.
4. TC% (V) (0,854): Esta variable representa el porcentaje de acierto en los intentos de campo del equipo visitante. Un valor alto en RC2 indica que el equipo visitante tiene una buena eficiencia en sus intentos de campo en general.
5. AST (V) (0,480): Esta variable refleja las asistencias realizadas por el equipo visitante. Un valor alto en RC2 sugiere que el equipo visitante tiene una capacidad destacada para generar jugadas de equipo y asistir a sus compañeros de manera efectiva.
6. ROB (V) (0,134): Esta variable se refiere a los robos de balón realizados por el equipo visitante. Un valor alto en RC2 indica que el equipo visitante tiene una habilidad notable para recuperar el balón mediante robos.

Además, la componente RC2 muestra correlaciones positivas con las variables AST (V) (0,480) y AST/PER (V) (0,254), lo que sugiere que el equipo también se focaliza en la generación de asistencias y en el juego colectivo para crear oportunidades de anotación para sus compañeros de equipo.

En resumen, la estrategia asociada con la componente RC2 sugiere que el equipo visitante tiende a realizar muchos intentos de lanzamiento de dos puntos y tiene una alta eficacia en el acierto de ese tipo de tiros. Además, el equipo visitante muestra una propensión a realizar numerosos

intentos de campo y tiene una buena eficiencia general en sus tiros. El equipo también se destaca en la generación de asistencias y en la habilidad para recuperar el balón mediante robos.

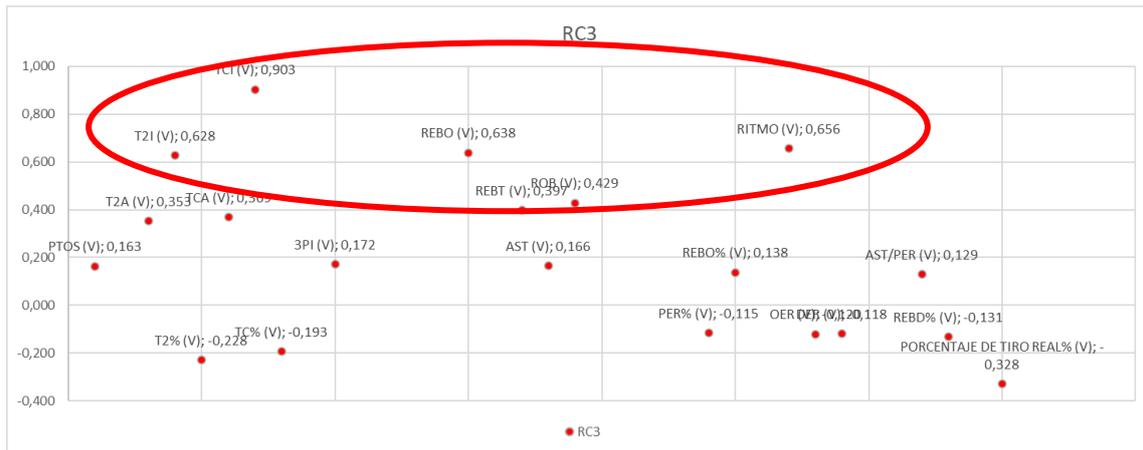


**Figura 38:** Gráfico componente RC2.

La componente RC3 presenta las siguientes variables con los mayores pesos:

1. TCI (V) (0,903): Representa el porcentaje de acierto en los intentos de campo del equipo visitante. Un valor alto indica que el equipo visitante tiene una buena eficiencia en sus intentos de campo en general.
2. REBO (V) (0,638): Refleja los rebotes ofensivos capturados por el equipo visitante. Un valor alto indica que el equipo visitante tiene una habilidad notable para obtener rebotes ofensivos.
3. RITMO (V) (0,656): Esta variable se refiere al ritmo de juego del equipo, que se basa en la velocidad y la frecuencia de los ataques. Un valor alto en RC3 sugiere que el equipo juega a un ritmo rápido, lo que indica una estrategia ofensiva orientada a correr la cancha, atacar rápidamente y generar transiciones rápidas.
4. ROB (V) (0,429): Representa los robos de balón realizados por el equipo visitante. Un valor alto indica que el equipo visitante tiene una habilidad destacada para recuperar el balón mediante robos.

En resumen, la estrategia asociada con la componente RC3 sugiere que el equipo visitante tiene una capacidad destacada para anotar puntos y muestra eficacia en los lanzamientos de dos puntos. El equipo también tiende a realizar muchos intentos de campo y tiene una buena eficiencia en sus tiros en general. Además, el equipo muestra habilidades en la captura de rebotes ofensivos, la generación de asistencias y la capacidad para recuperar el balón mediante robos.

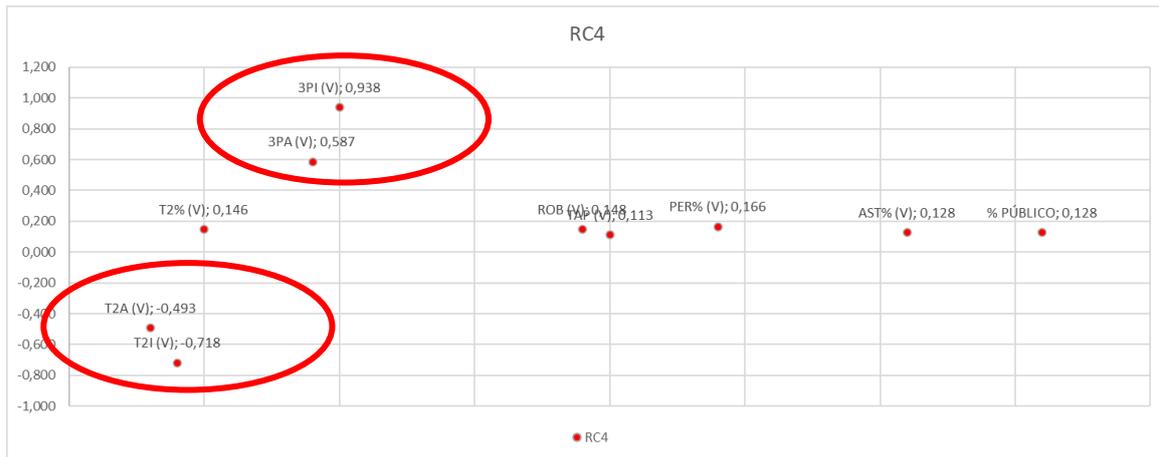


**Figura 39:** Gráfico componente RC3.

La componente RC4 presenta las siguientes variables con los mayores pesos:

1. T2A (V) (-0,493): Esta variable representa los intentos de lanzamiento de dos puntos del equipo visitante. Un valor bajo en RC4 sugiere que el equipo visitante tiende a realizar menos intentos de lanzamiento de dos puntos en sus partidos.
2. T2I (V) (-0,718): Esta variable se refiere al porcentaje de intentos de lanzamiento de dos puntos convertidos por el equipo visitante. Un valor bajo en RC4 indica que el equipo visitante tiene una eficacia relativamente baja en sus tiros de dos puntos.
3. 3PA (V) (0,587): Esta variable representa los intentos de lanzamiento de tres puntos del equipo visitante. Un valor alto en RC4 sugiere que el equipo visitante tiende a realizar muchos intentos de lanzamiento de tres puntos en sus partidos.
4. 3PI (V) (0,938): Esta variable se refiere al porcentaje de intentos de lanzamiento de tres puntos convertidos por el equipo visitante. Un valor alto en RC4 indica que el equipo visitante tiene una alta eficacia en sus tiros de tres puntos.

En resumen, la estrategia asociada con la componente RC4 implica una menor cantidad de intentos de tiros de dos puntos, lo que sugiere una preferencia por lo visto en las componentes de los lanzamientos de tres puntos. Además, el equipo muestra habilidades defensivas en términos de robos y puede beneficiarse de un alto porcentaje de asistencia de público en sus partidos.



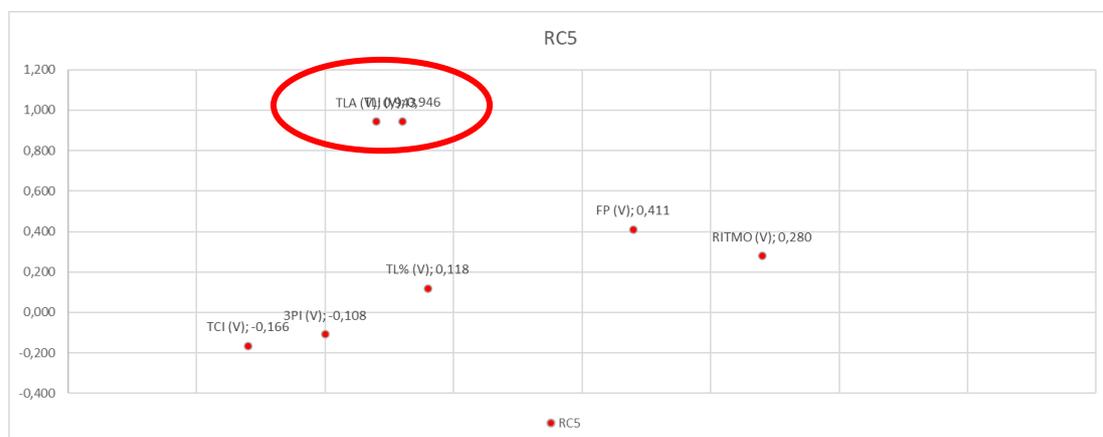
**Figura 40:** Gráfico componente RC4.

La componente RC5 presenta las siguientes variables con los mayores pesos:

1. TLA (V) (0,943): Esta variable representa los intentos de lanzamiento de tiros libres del equipo visitante. Un valor alto en RC5 indica que el equipo visitante tiende a recibir muchas oportunidades de tiros libres en sus partidos.
2. TLI (V) (0,946): Esta variable se refiere al porcentaje de tiros libres convertidos por el equipo visitante. Un valor alto en RC5 sugiere que el equipo visitante tiene una alta eficacia en los tiros libres.

Además, la componente RC5 muestra una correlación positiva con la variable RITMO (V) (0,280), lo que puede indicar que el equipo juega a un ritmo más acelerado, buscando generar transiciones rápidas y oportunidades de anotación.

En resumen, la estrategia asociada con la componente RC5 sugiere que el equipo visitante puede tener dificultades en la eficacia de sus tiros de campo y en los tiros de tres puntos. Sin embargo, el equipo destaca en los tiros libres, tanto en cantidad como en eficacia. El equipo también tiende a cometer un número significativo de faltas personales durante sus partidos. Por lo que se caracteriza de jugar un juego duro y con mucho contacto para forzar la mayor cantidad de faltas personales.



**Figura 41:** Gráfica componente RC5.

En este caso, la componente RC5, no es una estrategia real de juego en el baloncesto, ya que no hay demasiados tiros libres por partido como para enfocar una estrategia a esto. Al igual que hay faltas limitadas por lo que no te permite llevar a cabo esta estrategia, por lo que se decide eliminar.

Por tanto, tras ver los distintos componentes uno a uno, obtendríamos las siguientes estrategias:

	<i>Variables</i>	<i>Estrategia</i>
<b>Comp.1</b>	PTS (V), Porcentaje de tiro real% (V), Porcentaje Efectivo de Tiro% (V), OER (V)	Eficacia en la anotación y la eficiencia ofensiva y mantener la distancia en el marcador.
<b>Comp.2</b>	T2A (V), T2% (V), TCA (V), TC% (V)	Muchos lanzamientos de 2P, con una eficacia general buena, pero muy cambiantes.
<b>Comp.3</b>	TCI (V), RITMO (V), REBO (V), T2I (V)	Gran ritmo de juego, aprovechando las oportunidades en la transición y generar puntos rápidos y cargar fuerte el rebote ofensivo.
<b>Comp.4</b>	T2I (V), T2A (V), 3PI (V), 3PA (V)	Estrategia enfocada en el lanzamiento de 3 puntos.

*Tabla 14:* Estrategias seguidas cada componente.

### 3.1.2. Análisis Clúster

Tras identificar cada una de las estrategias de los equipos tanto como locales como visitantes por medio del PCA, se procedió a clasificar las variables en clústeres para así poder ver el éxito que han tenido las estrategias que se han definido anteriormente.

Para realizar este análisis, se utilizarán las componentes principales vistas en los apartados anteriores, en las estrategias como local y visitante. Nos quedaremos con las primeras 4 componentes en las estrategias como local y con las 5 primeras componentes en las estrategias como visitante.

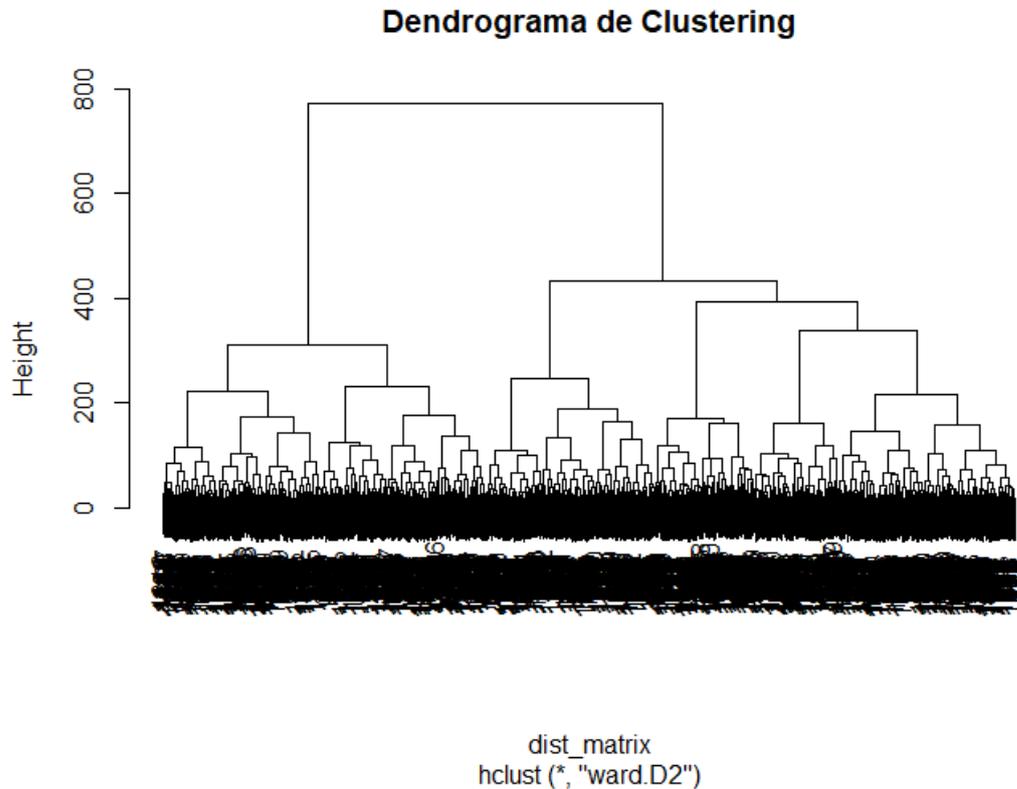
Al final de este análisis, se conseguirá tener la relación entre los partidos ganados y perdidos por cada uno de los clústeres, por lo que se podrá ver el éxito que ha tenido cada una de las estrategias.

#### 3.1.2.1. Análisis clúster jerárquico

##### 3.1.2.1.1. Equipos como local

Para hacer el análisis clúster jerárquico para los equipos que juegan como local, en este caso utilizaremos el método de Ward. El método de Ward es un algoritmo utilizado en el análisis de conglomerados para agrupar observaciones de manera que se minimice la varianza dentro de cada grupo. Se basa en calcular distancias euclídeas entre las observaciones y fusionar los grupos de manera iterativa minimizando la varianza. El resultado se representa en un dendrograma, que

muestra cómo se fusionaron los grupos. El método de Ward forma grupos compactos y homogéneos, pero puede ser sensible a valores atípicos y datos ruidosos.

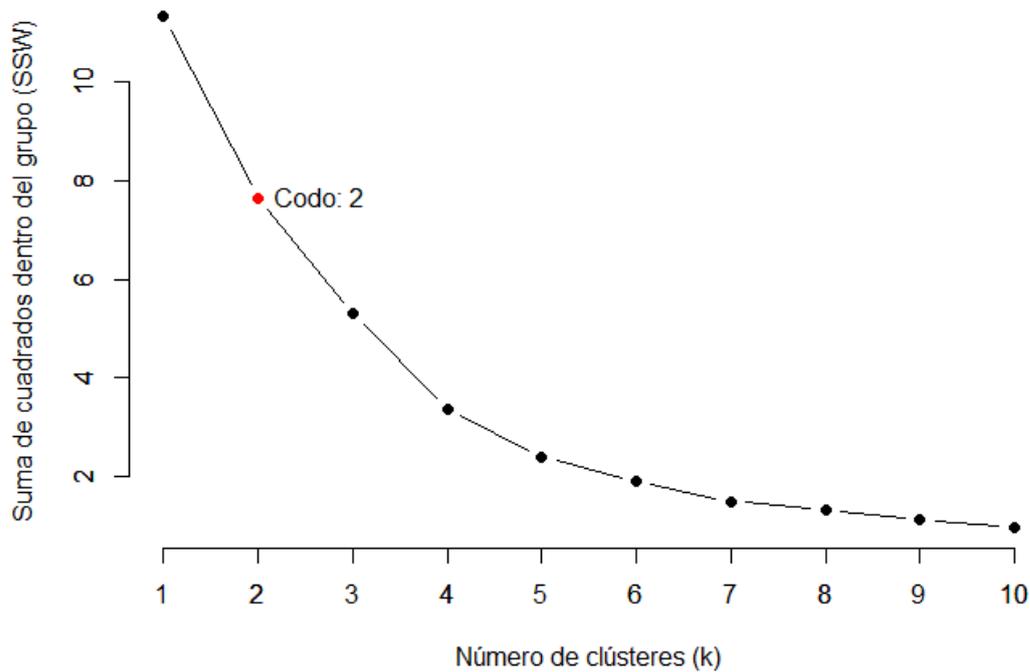


*Figura 42:* Dendrograma.

Para determinar el número óptimo de clústeres a partir del dendrograma vamos a utilizar el algoritmo k-means para evaluar el valor a partir del cual ya no hay una gran reducción de la suma de cuadrados intra-clúster y trabaja mediante el método del “codo”. El método del codo, también conocido como método de la suma de cuadrados dentro del grupo (SSD), es una técnica utilizada para determinar el número óptimo de clústeres. Se basa en la idea de que a medida que aumentamos el número de clústeres, la variabilidad dentro de cada clúster disminuye. El método del codo traza la suma de los cuadrados de las distancias de cada punto al centroide de su clúster correspondiente en función del número de clústeres. El número óptimo de clústeres se encuentra en el "codo" de la curva, donde el aumento en la reducción de la variabilidad se vuelve menos significativo.

Al interpretar el método del codo, se busca un punto en la curva donde agregar más clústeres ya no proporciona una mejora sustancial en la varianza explicada o la calidad de la agrupación en el análisis de clustering para ayudar en la interpretación y selección de clústeres adecuados.

Tras implementar el método del “codo” obtenemos el siguiente gráfico:



*Figura 43:* Método del "codo".

Para hacer una mejor comprobación, sacamos los registros totales de los clústeres para ver cómo están de balanceados, y obtenemos los siguientes resultados:

	Solución con 2 clústeres	Solución con 3 clústeres	Solución con 4 clústeres
<b>Clúster 1</b>	639	442	364
<b>Clúster 2</b>	442	407	375
<b>Clúster 3</b>		474	232
<b>Clúster 4</b>			352

*Tabla 15:* Balanceo de los clústeres.

Tras ver esto, vemos que la solución con dos, tres y cuatro clústeres son balanceados, no se encuentran grandes diferencias. En este caso, vamos a escoger 3 clústeres que parece el más balanceado de los tres y se procede analizar las estrategias predominantes en cada uno de ellos y para ello analizaremos los scores de cada uno de los clústeres sobre cada uno de los componentes obtenidos que se muestran en la Tabla 16.

Clúster	RC1	RC2	RC3	RC4
1	0,714	-0,018	0,124	-0,027
2	-0.056	0.695	0.097	-0.108
3	0.029	0.015	-0.138	0.205

*Tabla 16:* Scores obtenidos.

Tras obtener estos resultados, se puede describir las distintas estrategias de los dos clústeres que se han obtenido:

El clúster 1 se caracteriza por tener un valor más alto, que a su vez es el mayor valor de los tres clústeres, en la componente uno. Por lo tanto, se caracteriza por la eficacia en la anotación y la eficiencia ofensiva.

También se puede ver que hay valores negativos tanto en la componente 2 como en la componente 4, lo que podría indicar que la estrategia viene enfocada en la anotación y estas dos componentes se centra más en la búsqueda del rebote y las segundas oportunidades, y también el juego duro para buscar los lanzamientos de tiros libres. Pero, no tanto, la eficacia en la anotación, ya que estas dos componentes en relación con el juego no indica eficacia y eficiencia en el ataque, sino todo lo contrario.

Por otro lado, el clúster 2, tiene valor positivo en la componente 2, que indica una estrategia basada en el rebote, búsqueda de segundas oportunidades, ya que es posible que el equipo sea menos eficaz en el juego ofensivo, lo que cuadraría con que la componente 3 sea negativa que es una estrategia basada en la defensa.

Mientras como se puede ver, no tiene apenas relación con la componente 1, que es prácticamente 0, pero si relación con la componente 4, que tampoco muestra gran eficacia ofensiva.

Y, por último, el clúster 3 tendría como mayor peso la componente 4, que se basaría en una estrategia ofensiva basada en el aprovechamiento de los tiros libres.

Y, por último, vemos el éxito general de las estrategias, que, tras implementar el código, tenemos lo siguiente:

Clúster	Victoria	Derrota
1	0,438	0.563
2	0.667	0.333
3	0.643	0.357

*Tabla 17:* Proporción V/D obtenidos según estrategia.

Los resultados presentados muestran la proporción de partidos ganados y perdidos por clúster, así como el resultado de la prueba de chi-cuadrado realizada.

En la tabla de proporciones, se muestran tres clústeres identificados como "Clúster 1", "Clúster 2" y "Clúster 3". Para el "Clúster 1", la proporción de partidos ganados es del 43.8% y la proporción de partidos perdidos es del 56.3%. Para el "Clúster 2", la proporción de partidos ganados es del 66.7% y la proporción de partidos perdidos es del 33.3%. Mientras que el último clúster, la probabilidad de victoria es del 64.3%, mientras que de derrota es del 35.7%

Además, se realizó la prueba de chi-cuadrado de Pearson a las proporciones de V/D que obtuvimos en la tabla y se obtuvo lo siguiente:

```
> cat("Prueba de Chi-cuadrado:\n")
Prueba de Chi-cuadrado:
> print(chi_test)

      Pearson's Chi-squared test

data:  df_resultados_clusters$Resultados and df_resultados_clusters$cluster
X-squared = 1.6353, df = 2, p-value = 0.4415
```

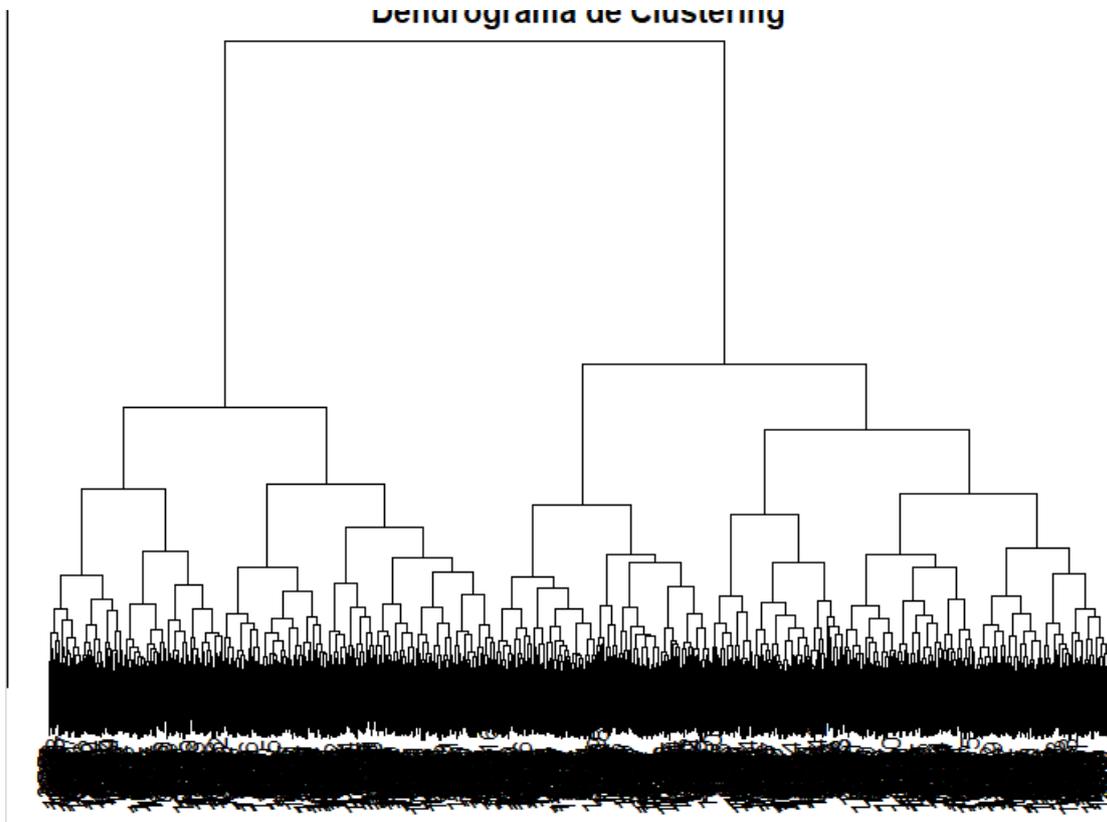
*Figura 44:* Prueba de chi-cuadrado.

En cuanto a la prueba de chi-cuadrado, el estadístico de chi-cuadrado calculado es de 1.6353 con 2 grados de libertad. El valor p obtenido es 0.4415. La prueba de chi-cuadrado se utiliza para determinar si hay una asociación significativa entre las variables "Resultados" (victoria/derrota) y "Clúster". En este caso, como el valor p (0.4415) es mayor que el nivel de significancia comúnmente utilizado (por ejemplo, 0.05), no hay suficiente evidencia para rechazar la hipótesis nula de que no hay asociación entre las variables (igualdad de proporciones).

En resumen, los resultados indican que no hay diferencias significativas en las proporciones de partidos ganados y perdidos entre los diferentes clústeres, según la prueba de chi-cuadrado. A nivel descriptivo podemos ver que el clúster 2 es el más exitoso en victorias y se caracteriza por tener una estrategia basada en el rebote, búsqueda de segundas oportunidades, ya que es posible que el equipo sea menos eficaz en el juego ofensivo. No obstante, al no ser significativo la diferencia indicaría que no existe una estrategia concreta de éxito o fracaso y que la victoria o derrota dependerá de los equipos sepan adaptarse a las distintas situaciones.

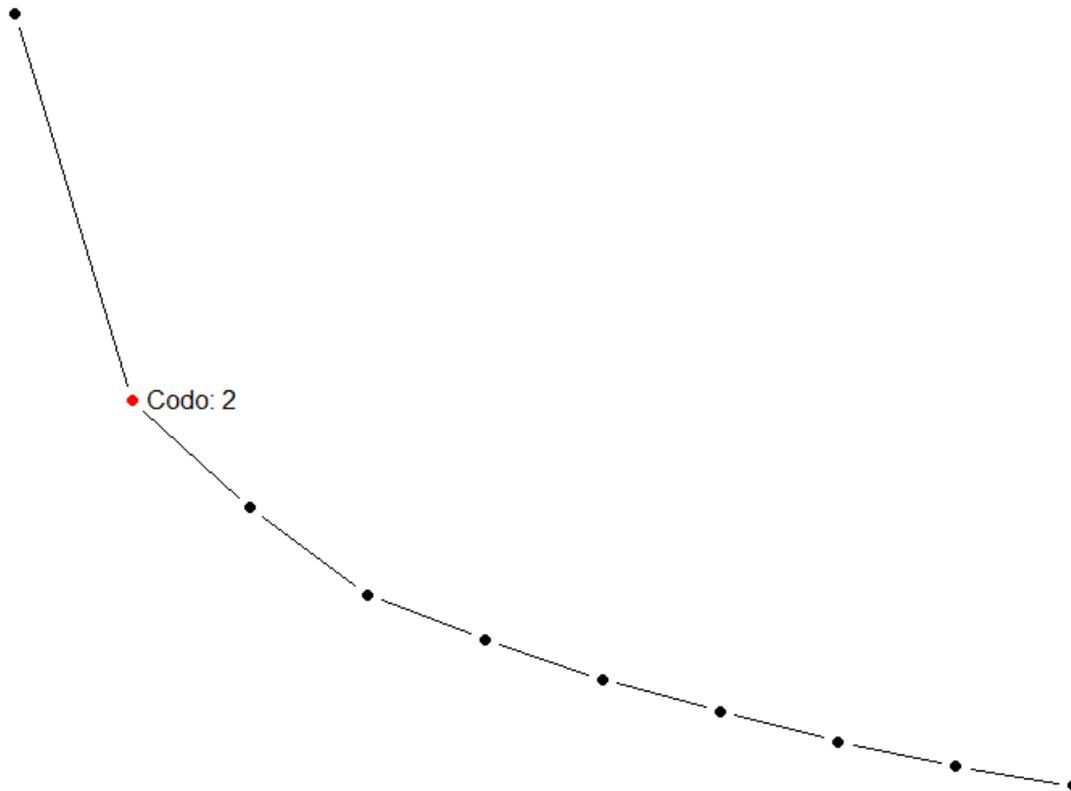
### 3.1.2.1.2. Equipos como visitante

Siguiendo los mismos pasos que en el apartado anterior, realizamos el dendrograma para los equipos actuando como visitante utilizando el método de Ward:



*Figura 45:* Dendrograma.

Tras esto, como se ha realizado anteriormente, se utilizará el algoritmo k-means que trabaja mediante el método del “codo” para escoger el número de clústeres adecuado y, tras esto, veremos los scores y definiremos las estrategias seguidos en cada uno de los clústeres:



*Figura 46:* Método del "codo".

Para hacer una mejor comprobación, sacamos los registros totales de los clústeres para ver cómo están de balanceados, y obtenemos los siguientes resultados:

	<b>Solución con 2 clústeres</b>	<b>Solución con 3 clústeres</b>	<b>Solución con 4 clústeres</b>
<b>Clúster 1</b>	688	319	343
<b>Clúster 2</b>	635	592	356
<b>Clúster 3</b>		412	272
<b>Clúster 4</b>			352

*Tabla 18:* Balanceo de los clústeres.

Tras ver esto, vemos que el número de clústeres a utilizar son dos, ya que con 3 y 4 los clústeres se encuentran más desbalanceados.

Tras ver esto, vemos que el número de clústeres a utilizar son dos. Por lo que para obtener las estrategias obtenidas en los dos clústeres obtenemos los scores y tenemos lo siguiente:

Clúster	RC1	RC2	RC3	RC4
1	0,823	0,027	-0,044	-0,011
2	-0.005	0.151	0.133	0.036

*Tabla 19:* Scores obtenidos.

Tras obtener estos resultados, se puede describir las distintas estrategias de los dos clústeres que se han obtenido:

El clúster 1 se caracteriza por tener un valor más alto, que a su vez es el mayor valor de los dos clústeres, en la componente uno. Por lo tanto, se caracteriza por la eficacia en la anotación y la eficiencia ofensiva y mantener la distancia en el marcador. También se puede ver que hay valores negativos tanto en la componente 3 y 4, esto puede ser visto como que la estrategia viene enfocada en la anotación y estas tres componentes se centra más en la búsqueda del rebote y las segundas oportunidades, y también el juego duro para buscar los lanzamientos de tiros libres, y además del juego rápido buscando un juego más rápido y con muchos lanzamientos. Pero, no tanto, la eficacia en la anotación, ya que estas dos componentes en relación con el juego no indica eficacia y eficiencia en el ataque, sino todo lo contrario.

Por otro lado, el clúster 2 se caracteriza por tener un valor más alto en la componente 2 y 3, lo que indica una estrategia basada en el lanzamiento de 2P, y cuadra con la componente 3 que sea ligeramente parecida, ya que se basa en tener un ritmo alto de juego.

Y, por último, vemos el éxito general de las estrategias, que, tras implementar el código, tenemos lo siguiente:

Clúster	Victoria	Derrota
1	0,167	0,833
2	0,500	0,500

*Tabla 20:* Proporción V/D obtenidos según estrategia.

Los resultados presentados muestran la proporción de partidos ganados y perdidos por clúster, así como el resultado de la prueba de chi-cuadrado realizada.

En la tabla de proporciones, se muestran cuatro clústeres identificados como "Clúster 1" y "Clúster 2". Para el "Clúster 1", la proporción de partidos ganados es del 16,7% y la proporción de partidos perdidos también es del 83,3%. Para el "Clúster 2", la proporción de partidos ganados es del 50% y la proporción de partidos perdidos es del 50%.

Además, se realizó la prueba de chi-cuadrado de Pearson a las proporciones de V/D que obtuvimos en la tabla y se obtuvo lo siguiente:

```
> cat("Prueba de Chi-cuadrado:\n")
Prueba de Chi-cuadrado:
> print(chi_test)

Pearson's Chi-squared test with Yates' continuity correction

data: df_resultados_clusters$Resultados and df_resultados_clusters$Cluster
X-squared = 1.1025, df = 1, p-value = 0.2937
```

*Figura 47:* Prueba de chi-cuadrado.

En cuanto a la prueba de chi-cuadrado, el estadístico de chi-cuadrado calculado es de 1.1025 con 1 grado de libertad. El valor p obtenido es 0.2937. La prueba de chi-cuadrado se utiliza para determinar si hay una asociación significativa entre las variables "Resultados" (victoria/derrota) y "Clúster". En este caso, como el valor p (0.2937) es mayor que el nivel de significancia comúnmente utilizado (por ejemplo, 0.05), no hay suficiente evidencia para rechazar la hipótesis nula de que no hay asociación entre las variables (proporciones distintas), pudiéndose concluir lo mismo que los equipos locales.

En resumen, los resultados indican que no hay diferencias significativas en las proporciones de partidos ganados y perdidos entre los diferentes clústeres, según la prueba de chi-cuadrado.

### 3.1.3. Análisis discriminante

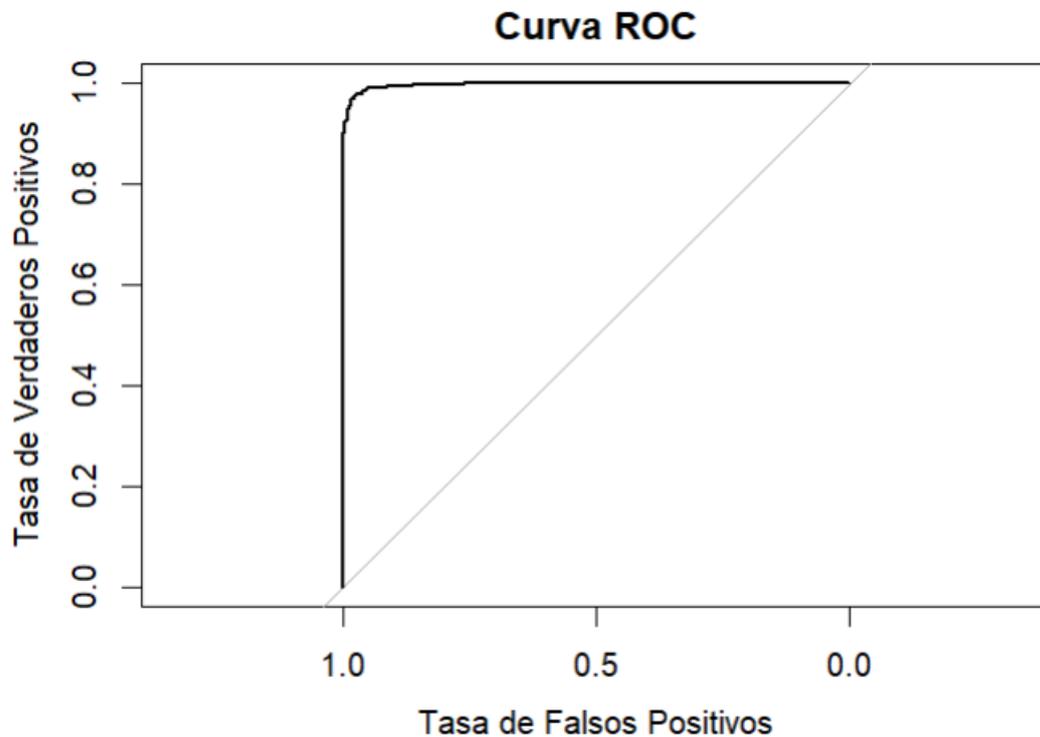
El análisis discriminante es una técnica estadística utilizada para clasificar y predecir el resultado de los partidos de baloncesto basándose en diferentes variables. En este caso, utilizaremos el análisis discriminante para analizar el éxito en la predicción de los partidos y determinar qué variables tienen un mayor peso en las predicciones.

#### 1.3.2.1 Análisis discriminante equipos como local

Primero, aplicamos el análisis discriminante utilizando la función `lda()` en RStudio. Esto nos proporciona un modelo discriminante que nos permite analizar tanto el éxito en la predicción como los coeficientes lineales de las diferentes variables.

En primer lugar, se analiza la curva ROC (Receiver Operating Characteristic) que consiste en una representación gráfica que muestra la capacidad de un modelo para distinguir entre dos categorías. En el contexto del análisis discriminante aplicado al baloncesto, la curva ROC nos permite evaluar lo bien que nuestro modelo puede predecir si un equipo ganará o perderá un partido, proporcionando una medida de su precisión global.

En nuestro caso, se obtiene la siguiente curva ROC:



*Figura 48:* Curva ROC.

Aquí, tras verlo en RStudio, se identifica que el área debajo de la curva es del 0.9975, lo que indica que la posibilidad de que clasifique correctamente el partido ganado y el partido perdido es del 99,75%.

La matriz de confusión con el conjunto de datos de prueba que muestra los siguientes resultados:

		Valores reales	
		No ganado	Ganado
Valores predichos	No ganado	581	15
	Ganado	17	710

*Tabla 21:* Matriz de confusión.

Con estos resultados, tenemos un 97,58% de aciertos. Se acierta cuando es un partido ganado el 99% de las ocasiones y cuando es un partido perdido también el 99% de las veces. Por lo tanto, podemos decir que son resultados especialmente buenos.

Por otro lado, obtenemos el siguiente peso de las variables en las predicciones:

`PTS (L)`	0.038581802
`Diferencia Marcador (L)`	-0.022832397
`T2A (L)`	0.024317102
`T2I (L)`	0.002409008
`T2% (L)`	-3.100522603
`3PA (L)`	-0.154059443
`3PI (L)`	0.025782113
`3P% (L)`	0.897307180
`TCA (L)`	-0.054200739
`TCI (L)`	0.025773177
`TC% (L)`	0.441970864
`TLA (L)`	0.258670590
`TLI (L)`	-0.055387342
`TL% (L)`	-1.485996588
`REBO (L)`	0.070664048
`REBD (L)`	0.003708963
`REBT (L)`	0.027611605
`AST (L)`	-0.160120274
`ROB (L)`	0.018777446
`TAP (L)`	0.008417212
`PER (L)`	-0.034033224
`FP (L)`	-0.036324463
`PORCENTAJE EFECTIVO DE TIRO% (L)`	0.156540866
`PER% (L)`	-0.032698835
`REBO% (L)`	-0.018618511
`TL% (EST AVAN - L)`	-0.148193451
`RITMO (L)`	-0.054233492
`OER (L)`	-1.670653985
`DER (L)`	-6.612956524
`NET RATING (L)`	6.037451432
`AST% (L)`	6.799571539
`AST/PER (L)`	-0.074496291
`REBD% (L)`	0.001560874
`REB% (L)`	-0.016768915
`PORCENTAJE DE TIRO REAL% (L)`	0.109282851
`% PÚBLICO`	0.180268101

*Figura 49:* Peso de las variables en las predicciones.

A partir de los coeficientes discriminantes lineales de cada una de las variables, aquellas que mayor peso tengan más peso tendrán en las predicciones, por lo tanto, podemos concluir que: las variables más influyentes son:

1. **NET RATING (L):** Peso positivo de 6.037451432.
2. **AST% (L):** Peso positivo de 6.799571539.
3. **PORCENTAJE DE TIRO REAL% (L):** Peso positivo de 0.109282851.
4. **% PÚBLICO:** Peso positivo de 0.180268101.
5. **TLA (L):** Peso positivo de 0.258670590

Estas variables son las que tienen mayor influencia positiva en la predicción del éxito en los partidos. Se pueden considerar como puntos clave para desarrollar una estrategia enfocada en maximizar el rendimiento del equipo. Esto se corresponde con una de las estrategias vistas anteriormente, en concreto, la siguiente:

	<i>Variables</i>	<i>Estrategia</i>
<b>Comp.1</b>	Porcentaje efectivo de tiro % (L), TC% (L), OER (L), PTS (L)	Eficacia en la anotación y la eficiencia ofensiva.

*Tabla 22:* Estrategia definida.

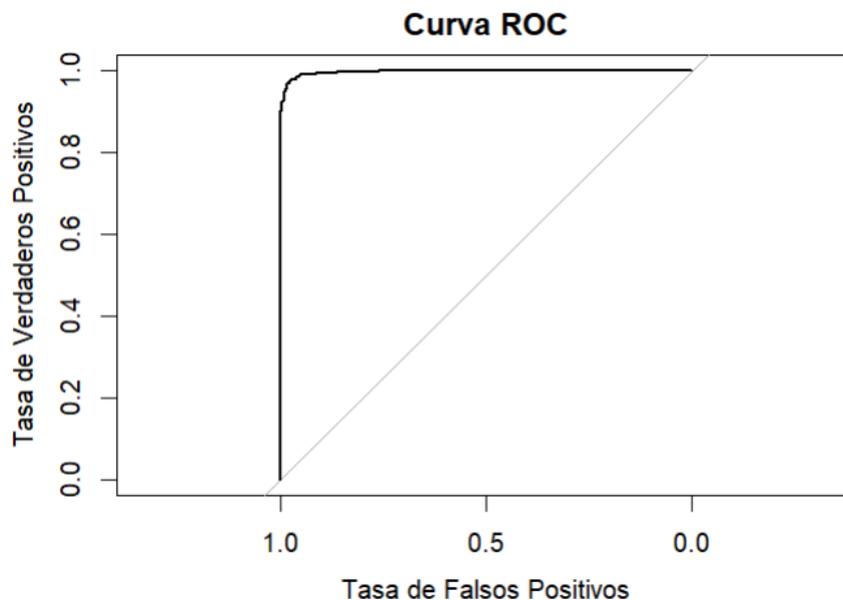
Las 6 variables con mayor peso negativo serán las que influyen en los partidos no ganados y son las siguientes:

1. **DER (L):** Peso negativo de -6.612956524.
2. **OER (L):** Peso negativo de -1.670653985.
3. **TL% (L):** Peso negativo de -1.485996588.
4. **T2% (L):** Peso negativo de -3.100522603.
5. **AST (L):** Peso negativo de -0.160120274.
6. **3PA (L):** Peso negativo de -0.154059443.

Estas variables tienen mayor influencia negativa en la predicción del éxito en los partidos. Se pueden considerar como áreas de mejora para desarrollar una estrategia enfocada en minimizar su impacto y mejorar el rendimiento del equipo. En este caso, no podemos definir, ninguna estrategia de las vistas anteriormente.

### 1.3.2.2 Análisis discriminante equipos como visitante

En este caso, se obtiene la siguiente curva ROC:



*Figura 50:* Curva ROC.

Aquí, tras verlo en RStudio, se identifica que el área debajo de la curva es del 0.999, lo que indica que la posibilidad de que clasifique correctamente el partido ganado y el partido perdido es del 99,9%.

La matriz de confusión que muestra los siguientes resultados:

		Valores reales	
		No ganado	Ganado
Valores predichos	No ganado	714	14
	Ganado	11	584

Tabla 23: Matriz de confusión.

Con estos resultados, tenemos un 98.11% de aciertos. Se acierta cuando es un partido ganado el 99% de las ocasiones y cuando es un partido perdido también el 99% de las veces. Por lo tanto, podemos decir que son resultados especialmente buenos.

Por otro lado, obtenemos el siguiente peso de las variables en las predicciones:

	LD1
`PTOS (V)`	-0.020447937
`Diferencia Marcador (V)`	0.081598918
`T2A (V)`	-0.043404271
`T2I (V)`	0.018601179
`T2% (V)`	0.501822519
`TCA (V)`	-0.029617702
`TCI (V)`	0.016756003
`TC% (V)`	5.564867154
`3PA (V)`	0.024275451
`3PI (V)`	-0.010087434
`3P% (V)`	-0.481668893
`TLA (V)`	-0.033885204
`TLI (V)`	0.023925988
`TL% (V)`	0.526958896
`REBO (V)`	0.031366963
`REBD (V)`	-0.011071848
`REBT (V)`	0.003758741
`AST (V)`	-0.024350930
`ROB (V)`	0.019642480
`TAP (V)`	-0.015254220
`PER (V)`	-0.026837552
`FP (V)`	-0.008431967
`PORCENTAJE EFECTIVO DE TIRO% (V)`	0.012231766
`PER% (V)`	0.003285329
`REBO% (V)`	0.006957522
`TL% (EST AVANZ - V)`	0.025118238
`RITMO (V)`	0.015412497
`OER (V)`	-0.500584446
`DER (V)`	-1.358890066
`NET RATING (V)`	1.431449788
`AST% (V)`	0.827610008
`AST/PER (V)`	0.095534250
`REBD% (V)`	-0.010462239
`REB% (V)`	-0.003952632
`PORCENTAJE DE TIRO REAL% (V)`	0.061961813
`% PÚBLICO`	-0.205028829

Figura 51: Peso de las variables en las predicciones.



Atendiendo de nuevo al peso de los coeficientes de cada variable, las 6 variables con mayor peso positivo son:

1. **TC% (V):** Peso positivo de 5.564867154.
2. **NET RATING (V):** Peso positivo de 1.431449788.
3. **AST% (V):** Peso positivo de 0.827610008.
4. **Diferencia Marcador (V):** Peso positivo de 0.081598918.
5. **TL% (V):** Peso positivo de 0.526958896.
6. **PORCENTAJE DE TIRO REAL% (V):** Peso positivo de 0.061961813.

Estas variables son las que tienen mayor influencia positiva en la predicción del éxito en los partidos. Se pueden considerar como puntos clave para desarrollar una estrategia enfocada en maximizar el rendimiento del equipo. Esto no se corresponde con ninguna de las estrategias definidas.

Las 6 variables con mayor peso negativo son:

1. **DER (V):** Peso negativo de -1.358890066.
2. **OER (V):** Peso negativo de -0.500584446.
3. **3P% (V):** Peso negativo de -0.481668893.
4. **% PÚBLICO:** Peso negativo de -0.205028829.
5. **PER (V):** Peso negativo de -0.026837552.
6. **TLA (V):** Peso negativo de -0.033885204.

Estas variables tienen mayor influencia negativa en la predicción del éxito en los partidos. Se pueden considerar como áreas de mejora para desarrollar una estrategia enfocada en minimizar su impacto y mejorar el rendimiento del equipo. En este caso, no podemos definir, ninguna estrategia de las vistas anteriormente. Si que es cierto, que cuadra que el peso negativo, es decir las derrotas como equipo visitante se vean influenciadas por el público y las pérdidas.

### 3.2. Influencia en la clasificación de las estrategias

En los equipos como local, como hemos visto tenemos una estrategia definida que indica que la eficiencia y eficacia ofensiva ayuda a conseguir la victoria, por lo que, si lo corroboramos con la clasificación final y las estadísticas, filtrándolas por la eficiencia ofensiva (OER), obtenemos los siguientes datos:

TEAM	GP	W	L	MIN	OFFRTO	DEFRTO	NETRTO	AST%	AST/TO	AST RATIO	OREB%	DREB%	REB%	TOV%	FTG%	TS%	PACE	PIE	POSS
1 Atlanta Hawks	41	27	14	1968.0	117.7	112.7	5.0	60.1	2.15	18.5	26.1	72.9	50.2	11.8	55.8	59.8	99.12	53.2	4065
2 Chicago Bulls	41	27	14	1973.0	115.3	113.1	2.2	56.8	2.05	18.0	25.7	74.5	50.3	12.0	54.5	58.3	98.71	50.9	4062
3 Indiana Pacers	41	16	25	1983.0	114.8	115.5	-0.7	63.3	1.91	19.2	31.7	72.6	51.8	14.3	54.4	57.5	98.71	49.7	4081
4 Miami Heat	41	29	12	1983.0	114.5	108.7	5.8	64.3	1.81	19.0	28.0	73.4	51.1	14.7	55.2	59.0	96.02	53.6	3965
5 Toronto Raptors	41	24	17	1983.0	114.2	110.0	4.2	54.7	1.85	16.5	33.2	71.1	51.2	12.6	52.0	55.1	96.36	50.7	3984
6 Milwaukee Bucks	41	27	14	1968.0	114.1	110.2	3.9	57.4	1.86	17.4	26.4	75.3	51.3	12.7	54.3	57.8	100.91	52.0	4142
7 Boston Celtics	41	28	13	1983.0	113.2	106.2	7.0	60.2	1.79	18.0	28.7	72.6	51.1	14.0	53.8	57.5	97.07	54.4	4012
8 Philadelphia 76ers	41	24	17	1983.0	113.1	109.7	3.4	62.0	1.97	18.4	25.1	71.6	48.9	12.8	53.5	57.8	96.77	52.0	3995
9 Washington Wizards	41	21	20	1993.0	112.3	113.1	-0.7	62.0	1.99	18.8	25.0	73.2	49.8	13.0	54.0	57.7	98.34	51.4	4085
10 Charlotte Hornets	41	22	19	1993.0	112.3	110.8	1.5	66.7	2.02	19.7	28.3	71.3	49.8	13.8	53.8	56.4	99.84	51.0	4143
11 Cleveland Cavaliers	41	25	16	1978.0	110.9	106.7	4.2	64.1	1.75	18.8	27.9	73.3	51.1	14.9	54.0	57.2	96.84	52.5	3997
12 Brooklyn Nets	41	20	21	1983.0	110.3	112.1	-1.8	60.9	1.66	18.0	29.1	70.6	50.4	15.1	52.4	56.3	99.47	50.3	4106
13 New York Knicks	41	17	24	1983.0	108.5	110.5	-2.1	59.6	1.65	16.6	28.0	74.2	50.8	13.8	51.5	55.0	96.77	48.5	4007
14 Detroit Pistons	41	13	28	1983.0	106.1	111.5	-5.4	61.1	1.66	17.1	28.7	70.2	48.8	14.3	49.4	53.4	98.20	46.4	4060
15 Orlando Magic	41	12	29	1983.0	103.7	112.0	-8.3	61.8	1.65	17.4	24.5	74.2	49.0	14.3	50.2	53.7	99.67	45.9	4122

Figura 52: Estadística filtrada por eficiencia ofensiva.

TEAM	GP	W	L	WIN%	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	+/-
1 Miami Heat	41	29	12	.707	48.4	110.8	40.0	84.4	47.4	13.1	34.8	37.7	17.5	21.5	81.4	9.9	33.5	43.4	25.7	14.2	7.3	3.1	4.1	20.4	20.8	5.5
2 Boston Celtics	41	28	13	.683	48.4	110.8	40.7	87.3	46.6	12.6	35.7	35.2	16.8	20.5	81.8	10.9	34.7	45.6	24.5	13.7	7.5	5.7	4.4	18.0	19.2	7.0
3 Atlanta Hawks	41	27	14	.659	48.0	116.7	41.9	87.0	48.2	13.3	35.2	37.9	19.5	23.8	82.1	9.3	34.3	43.6	25.2	11.7	7.8	4.6	4.0	18.9	21.1	5.0
3 Chicago Bulls	41	27	14	.659	48.1	114.2	42.8	88.2	48.5	10.6	28.6	37.1	18.0	22.2	81.0	9.3	34.3	43.7	24.3	11.9	7.0	4.1	5.3	18.3	18.2	2.4
3 Milwaukee Bucks	41	27	14	.659	48.0	115.3	41.7	89.7	46.4	14.1	38.6	36.6	17.9	22.9	78.0	10.2	37.5	47.7	23.9	12.9	7.5	3.9	4.4	18.6	18.9	4.2
6 Cleveland Cavaliers	41	25	16	.610	48.2	108.1	39.6	84.7	46.8	12.2	33.9	36.0	16.6	22.2	74.7	10.2	34.6	44.8	25.4	14.5	7.3	4.2	4.1	17.5	20.0	4.4
7 Philadelphia 76ers	41	24	17	.585	48.4	110.2	39.7	85.1	46.6	11.7	32.7	35.8	19.1	23.1	82.8	8.6	33.5	42.1	24.6	12.5	7.9	5.5	4.7	19.2	19.1	3.1
7 Toronto Raptors	41	24	17	.585	48.4	111.0	41.3	91.2	45.3	12.2	34.7	35.3	16.2	21.6	75.0	13.5	32.0	45.4	22.6	12.2	9.0	4.8	4.4	20.3	19.1	4.2
9 Charlotte Hornets	41	22	19	.537	48.6	113.5	42.4	91.3	46.4	13.6	38.7	35.0	15.1	20.9	72.3	11.4	35.0	46.3	28.3	14.0	7.9	5.6	4.7	20.0	19.7	1.4
10 Washington Wizards	41	21	20	.512	48.6	111.9	41.5	86.8	47.8	10.8	31.4	34.3	18.1	23.0	78.8	9.0	34.5	43.5	25.7	12.9	6.3	5.1	3.8	19.1	20.8	-0.6
11 Brooklyn Nets	41	20	21	.488	48.4	110.4	41.2	88.4	46.6	10.3	31.1	33.2	17.7	21.9	80.9	11.0	34.5	45.5	25.1	15.1	7.0	5.7	4.6	19.7	19.7	-2.0
12 New York Knicks	41	17	24	.415	48.4	106.0	37.3	85.8	43.5	13.7	38.0	36.0	17.7	24.0	73.9	10.9	34.7	45.6	22.2	13.5	6.8	5.3	4.5	19.8	20.1	-1.5
13 Indiana Pacers	41	16	25	.390	48.4	114.2	43.0	90.3	47.6	12.3	35.4	34.8	15.9	20.5	77.6	11.6	32.2	43.8	27.2	14.2	7.4	5.5	4.8	20.1	18.8	-0.6
14 Detroit Pistons	41	13	28	.317	48.4	105.1	38.4	88.7	43.3	10.9	34.3	31.6	17.4	21.9	79.5	11.0	32.6	43.6	23.4	14.1	7.8	4.9	5.2	22.4	19.8	-5.1
15 Orlando Magic	41	12	29	.293	48.4	104.2	38.4	88.2	43.5	11.7	36.4	32.0	15.8	20.0	79.0	8.8	34.9	43.7	23.7	14.4	6.9	4.5	5.0	19.6	18.7	-8.1

Figura 53: Clasificación Conferencia Este.

Como se puede ver, de los 8 equipos que se clasificaron a play-off, seis se encuentran entre los ocho mejores equipos en cuanto a la eficiencia ofensiva. Solo dos de ellos, no estaría dentro de este grupo.

Si vemos ahora, el mismo caso, pero en la conferencia oeste, tenemos lo siguiente:

TEAM	GP	W	L	MIN	OFFRTG	DEFRTG	NETRTG	AST%	AST/TO	AST RATIO	OREB%	DREB%	REB%	TOV%	EFG%	TS%	PACE	PIE	POSS
1 Utah Jazz	41	29	12	1973.0	118.4	108.4	10.0	56.5	1.71	17.3	29.9	74.2	53.0	13.9	56.5	59.9	98.38	54.7	4045
2 Memphis Grizzlies	41	30	11	1973.0	116.3	107.4	8.9	60.2	2.10	18.6	32.3	73.9	52.7	12.6	53.7	56.4	100.56	54.8	4137
3 Denver Nuggets	41	23	18	1983.0	115.5	113.1	2.4	66.9	2.07	20.1	26.5	76.4	51.6	13.6	55.6	59.2	98.66	51.5	4076
4 Phoenix Suns	41	32	9	1968.0	115.3	106.4	8.9	63.6	2.12	19.8	27.1	72.9	51.0	13.1	55.7	58.7	100.20	55.7	4107
5 New Orleans Pelicans	41	19	22	1978.0	113.6	112.0	1.5	62.5	1.86	18.6	31.7	73.4	52.1	14.2	52.8	56.5	97.65	51.3	4021
6 Dallas Mavericks	41	29	12	1978.0	113.3	106.8	6.5	59.0	1.94	17.6	26.2	72.3	49.5	12.5	54.0	57.2	95.10	52.5	3921
7 San Antonio Spurs	41	16	25	1973.0	113.3	113.3	0.0	64.4	2.19	19.5	28.2	71.1	49.5	12.6	53.6	56.5	100.12	50.8	4113
8 Golden State Warriors	41	31	10	1978.0	113.2	103.6	9.7	67.9	1.89	20.0	27.3	73.2	51.3	15.0	56.0	58.7	99.19	55.9	4092
9 Minnesota Timberwolves	41	26	15	1973.0	111.7	106.7	4.9	63.2	1.78	18.0	28.0	71.3	49.5	14.1	52.8	56.4	101.30	52.1	4165
10 LA Clippers	41	25	16	1973.0	110.7	108.3	2.3	59.4	1.70	17.7	26.0	71.0	49.2	14.3	54.0	57.3	98.96	51.4	4070
11 Sacramento Kings	41	16	25	1973.0	110.2	114.1	-3.9	59.3	1.66	17.5	25.0	71.2	48.3	14.3	53.2	57.1	99.92	48.2	4104
12 Houston Rockets	41	11	30	1983.0	110.1	116.6	-6.5	61.6	1.56	17.8	25.7	70.3	47.9	15.5	54.3	57.3	101.00	46.2	4169
13 Los Angeles Lakers	41	21	20	1998.0	110.1	110.4	-0.4	57.9	1.67	17.5	26.0	71.2	48.9	14.3	53.7	56.6	100.64	49.7	4186
14 Portland Trail Blazers	41	17	24	1978.0	109.6	114.8	-5.2	59.1	1.65	17.1	27.8	72.6	49.5	14.1	52.1	55.6	98.68	46.7	4070
15 Oklahoma City Thunder	41	12	29	1973.0	104.9	111.4	-6.5	58.1	1.62	16.7	25.7	74.9	49.7	14.0	50.8	53.8	99.48	46.1	4090

Figura 54: Estadística filtrada por eficiencia ofensiva.

TEAM	GP	W	L	WIN%	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	+/-
1 Phoenix Suns	41	32	9	.780	48.0	115.5	44.0	89.7	49.0	12.0	32.3	37.3	15.5	19.7	78.6	10.2	36.1	46.2	28.0	13.2	8.0	4.9	3.9	19.2	18.4	8.9
2 Golden State Warriors	41	31	10	.756	48.2	113.0	41.5	87.6	47.4	15.0	40.7	36.9	15.0	19.5	76.6	10.1	36.6	46.8	28.2	14.9	9.1	4.8	3.6	20.8	17.1	9.9
3 Memphis Grizzlies	41	30	11	.732	48.1	117.3	44.6	94.4	47.2	12.2	33.3	36.7	16.0	21.8	73.3	13.3	35.8	49.0	26.8	12.8	9.7	6.8	5.7	19.0	19.1	9.2
4 Dallas Mavericks	41	29	12	.707	48.2	108.4	39.3	85.5	46.0	13.7	38.8	35.3	16.1	21.2	75.9	9.7	33.6	43.2	23.2	11.9	6.7	3.8	3.6	19.4	20.7	6.3
4 Utah Jazz	41	29	12	.707	48.1	116.8	41.6	86.7	48.0	14.9	40.8	36.4	18.8	24.7	76.2	10.5	36.6	47.1	23.5	13.7	6.8	5.3	3.9	19.5	20.8	9.9
6 Minnesota Timberwolves	41	26	15	.634	48.1	113.4	40.3	90.2	44.7	14.6	41.9	34.8	18.2	23.7	76.9	11.3	34.6	45.9	25.5	14.3	8.7	5.9	4.8	21.6	21.2	5.0
7 LA Clippers	41	25	16	.610	48.1	109.9	40.5	87.5	46.3	13.3	34.7	38.5	15.4	19.2	80.5	9.7	34.7	44.3	24.1	14.2	7.7	4.8	3.9	18.4	18.4	2.5
8 Denver Nuggets	41	23	18	.561	48.4	114.8	42.0	87.1	48.2	12.9	36.5	35.2	17.9	22.3	80.0	9.1	34.9	44.0	28.1	13.6	7.0	3.5	4.8	20.6	20.4	2.4
9 Los Angeles Lakers	41	21	20	.512	48.7	112.4	42.1	89.1	47.3	11.5	33.4	34.5	16.6	23.0	71.9	9.7	34.6	44.3	24.4	14.6	8.3	5.8	4.1	20.3	20.3	-0.5
10 New Orleans Pelicans	41	19	22	.463	48.2	111.4	41.4	88.9	46.5	11.1	33.4	33.1	17.6	22.0	79.8	12.1	33.2	45.3	25.9	13.9	8.1	4.0	4.1	19.8	20.0	1.3
11 Portland Trail Blazers	41	17	24	.415	48.2	108.8	39.2	88.1	44.5	13.3	37.8	35.2	17.0	21.9	77.7	10.7	32.9	43.6	23.2	14.0	8.1	4.8	5.0	20.9	19.4	-5.0
12 Sacramento Kings	41	16	25	.390	48.1	110.3	40.0	86.2	46.5	11.7	32.6	35.7	18.6	23.8	78.1	9.0	33.2	42.2	23.8	14.3	7.3	4.7	4.8	19.2	20.2	-4.0
12 San Antonio Spurs	41	16	25	.390	48.1	113.6	43.0	91.0	47.3	11.6	32.4	35.8	16.1	21.8	73.8	10.9	34.3	45.1	27.7	12.6	7.1	4.9	4.5	18.6	19.5	-0.1
14 Oklahoma City Thunder	41	12	29	.293	48.1	104.7	38.9	88.8	43.8	12.3	37.3	33.1	14.5	19.0	76.2	10.6	35.2	45.8	22.6	14.0	7.7	5.0	5.6	18.6	17.4	-6.4
15 Houston Rockets	41	11	30	.268	48.4	111.9	40.0	87.0	46.0	14.5	39.8	36.4	17.5	24.3	71.9	8.9	32.4	41.3	24.6	15.8	7.1	4.8	6.1	21.0	22.3	-6.8

Figura 55: Clasificación Conferencia Oeste.

Como se puede ver, en este caso también, de los 8 equipos que se clasificaron a play-off, seis se encuentran entre los ocho mejores equipos en cuanto a la eficiencia ofensiva. Solo dos de ellos, no estaría dentro de este grupo.

Respecto a las otras casuísticas, que serían, las derrotas en los equipos como local, y las victorias y derrotas como equipo visitante, no se ha podido definir una estrategia diferencial, por lo que no podemos corroborarlo en la clasificación final.

## Capítulo 4. Conclusiones y líneas futuras.

### 4.1. Conclusiones

En este proyecto se ha llevado a cabo distintos métodos estadísticos, en la que se ha realizado tanto un análisis no supervisado con la realización del análisis de componentes principales y el clúster, un análisis supervisado mediante la realización de un análisis discriminante, con la finalidad de determinar las principales estrategias de éxito en los partidos de baloncesto de la NBA

En primer lugar, realizamos un primer análisis descriptivo, donde hemos podido detectar la relación entre cada una de las variables, y sus características principales.

A continuación, para identificar las estrategias seguidas para los equipos, se procedió a diferenciar si jugaban de local como de visitante, determinando para cada una de las estrategias definidas las principales variables que las caracterizaba. También, ha sido posible comprobar cuáles son las variables que mayor peso han tenido en cada una de las estrategias, la cual nos ha ayudado a diferenciar cada estilo de juego.

Además, se ha podido determinar el éxito general que ha tenido cada una de las estrategias.

Por último, gracias a la realización del análisis discriminante se ha podido determinar cuál ha sido el éxito general de la predicción de las victorias y derrotas, de los equipos tanto como local, como visitante. Y, además, se ha podido ver cuál ha sido el peso de cada una de las variables en la predicción, lo que ha podido determinar la estrategia que favorece más a la victoria en los equipos como local. En cambio, no se ha podido definir para el resto de las casuísticas.

En definitiva, se han cumplido cada uno de los objetivos propuestos para alcanzar el objetivo final buscado en el proyecto: la identificación de las estrategias de los equipos de la NBA y su influencia en la clasificación durante una temporada regular.

### 4.2. Líneas Futuras

Como líneas futuras podemos distinguir las siguientes:

- Un análisis más profundo de los análisis supervisados y no supervisados, con más técnicas, que permitan conseguir unos resultados más exactos.
- Otra de las líneas futuras vendría relacionada con la realización de una interfaz, que permitiera la inclusión de una base de datos con estadísticas de cualquier liga y cualquier temporada de baloncesto, y que permitiera obtener cada una de las estrategias e influencias de manera automática.



## Bibliografía

- Airado-Rodríguez, D. ., (2017). REDUCCIÓN DE LA DIMENSIONALIDAD PARA EL ANÁLISIS E INTERPRETACION AFECTIVA EN LA PRÁCTICA EDUCATIVA. *Revista INFAD de Psicología " International Journal of Developmental and Educational Psychology*, págs. 97-102.
- basketanalysis.com. (s.f.). *basketanalysis.com*. Obtenido de [https://www.basketanalysis.com/index.php/sumario-de-la-estadistica-avanzada/#:~:text=Posesiones%20que%20finalizan%20en%20una,%2B%20FTA%20\\*%200%2C44](https://www.basketanalysis.com/index.php/sumario-de-la-estadistica-avanzada/#:~:text=Posesiones%20que%20finalizan%20en%20una,%2B%20FTA%20*%200%2C44))
- Bual, P. (6 de Mayo de 2012). *World Hoopstats*. Obtenido de <http://www.worldhoopstats.com/news/28-estadisticas-avanzadas.html>
- Calvo, D. (9 de Marzo de 2018). *DIEGO CALVO*. Obtenido de <https://www.diegocalvo.es/cluster-jerarquicos-y-no-jerarquicos/>
- Campus, B. D. (s.f.). *BIG DATA International Campus*. Obtenido de <https://www.campusbigdata.com/big-data-blog/item/116-big-data-en-el-deporte>
- Córdoba, U. d. (s.f.). *Universidad de Córdoba*. Obtenido de Universidad de Córdoba: [http://www.uco.es/zootecniaygestion/img/pictorex/09\\_13\\_25\\_sesion\\_8.pdf](http://www.uco.es/zootecniaygestion/img/pictorex/09_13_25_sesion_8.pdf)
- Datademia. (s.f.). *Datademia*. Obtenido de <https://datademia.es/blog/que-es-r>
- ESPN Enterprises, I. (s.f.). *ESPN DEPORTES*. Obtenido de ESPN DEPORTES: <https://espndeportes.espn.com/>
- Gallardo. (s.f.). *UGR*. Obtenido de UGR: <https://www.ugr.es/~gallardo/pdf/cluster-3.pdf>
- García de Zúñiga, F. (2 de Octubre de 2020). *arsys*. Obtenido de <https://www.arsys.es/blog/rstudio>
- González Ogando, P. (2019). Las estadísticas avanzadas en el baloncesto. *revistasuma*, 33-40.
- Grudemi, E. (Mayo de 2020). *Enciclopedia Económica*. Obtenido de <https://enciclopediaeconomica.com/variable-cualitativa/>
- Halweb. (s.f.). *Halweb*. Obtenido de Halweb: <https://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema3am.pdf>
- IBM. (7 de Diciembre de 2021). *IBM*. Obtenido de <https://www.ibm.com/docs/es/spss-statistics/beta?topic=features-discriminant-analysis>
- INESDI. (2 de Noviembre de 2022). Obtenido de <https://www.inesdi.com/blog/que-es-aprendizaje-no-supervisado/>



- inesfbl. (4 de Junio de 2015). *inesfbl*. Obtenido de <https://inesfbl.wordpress.com/2015/06/04/que-requisitos-necesito-para-poder-hacer-un-analisis-discriminante-y-cuales-los-pasos-a-seguir/>
- López Sánchez, J. M. (2021/2022). *Identificación de estrategias de juego en ligas europeas de fútbol*. Trabajo de Fin de Máster.
- Madrid, U. A. (s.f.). *estadistica.net*. Obtenido de estadistica.net: [https://www.estadistica.net/Master-Econometria/Analisis\\_Discriminante.pdf](https://www.estadistica.net/Master-Econometria/Analisis_Discriminante.pdf)
- matraco, & osopinoso. (s.f.). *Jugones*. Obtenido de <https://jugones.wordpress.com/2015/06/17/los-cuatro-factores-four-factors/>
- Monje, A. (28 de Noviembre de 2019). *GIGANTES*. Obtenido de [https://www.gigantes.com/nba/info-basica-nba/la-estadistica-avanzada-en-el-baloncesto-diccionario-de-conceptos-explicaciones-y-utilidades-por-andres-monje/#:~:text=Ritmo%20\(Pace\)%3A%20n%C3%BAmero%20de,ese%20jugador%20est%C3%A1%20en%20cancha.](https://www.gigantes.com/nba/info-basica-nba/la-estadistica-avanzada-en-el-baloncesto-diccionario-de-conceptos-explicaciones-y-utilidades-por-andres-monje/#:~:text=Ritmo%20(Pace)%3A%20n%C3%BAmero%20de,ese%20jugador%20est%C3%A1%20en%20cancha.)
- Moran Fanjul, J. (19 de Agosto de 2022). *hiberus*. Obtenido de <https://www.hiberus.com/crecemos-contigo/analisis-de-componentes-principales/>
- NBA. (s.f.). *NBA*. Obtenido de <https://www.nba.com/stats>
- Polonio Sánchez, M. (2021-2022). *Estudio del Análisis Discriminante. Aplicación a datos reales*. Trabajo de Fin de Grado.
- Prasniewski, A. (s.f.). *Researchgate*. Obtenido de [https://www.researchgate.net/figure/Figura-1-Analise-por-componentes-principais-Grafico-de-scores-a-e-grafico-de-loadings\\_fig1\\_329894151](https://www.researchgate.net/figure/Figura-1-Analise-por-componentes-principais-Grafico-de-scores-a-e-grafico-de-loadings_fig1_329894151)
- Rabinal, S. (16 de Octubre de 2022). *The Sporting News*. Obtenido de <https://www.sportingnews.com/es/nba/news/estadistica-avanzada-diccionario-terminos-que-es-que-sirve-nba-analytics-sabermetrics/1gc73ududvomy1s3mqiahg238c>
- Recuero De Los Santos, P. (2 de Diciembre de 2021). *Telefónica Tech*. Obtenido de <https://telefonicatech.com/blog/que-algoritmo-elegir-en-ml-aprendizaje>
- Riera, L. (22 de Octubre de 2017). *basketanalysis.com*. Obtenido de <https://www.basketanalysis.com/index.php/2017/10/22/los-four-factors-en-la-liga-endesa/>
- Rus Arias, E. (1 de Mayo de 2021). *Economipedia*. Obtenido de [https://economipedia.com/definiciones/analisis-cluster.html?nab=1&utm\\_referrer=https%3A%2F%2Fwww.google.com%2F](https://economipedia.com/definiciones/analisis-cluster.html?nab=1&utm_referrer=https%3A%2F%2Fwww.google.com%2F)
- Sandoval, W. (15 de Agosto de 2022). *rstudio-pubs-static*. Obtenido de [https://rstudio-pubs-static.s3.amazonaws.com/931871\\_33029808c35b48c3a2f598cf34ba0dad.html](https://rstudio-pubs-static.s3.amazonaws.com/931871_33029808c35b48c3a2f598cf34ba0dad.html)



Soporte, S. (Agosto de 2018). *Maxima Formacion*. Obtenido de <https://www.maximaformacion.es/blog-dat/que-es-r-software/>

Staff, N. (s.f.). *The Sporting News*. Obtenido de <https://www.sportingnews.com/ar/nba/news/que-estadios-nba-permiten-presencia-de-publico-temporada-2020-2021/1bdxofj8shr5p1s2xk5xg2wjku>

Tibco. (s.f.). *Tibco*. Obtenido de <https://www.tibco.com/es/reference-center/what-is-unsupervised-learning>

TODOBASKET. (s.f.). *TODOBASKET*. Obtenido de <https://todobasket.eu/publicaciones-de-baloncesto/estadistica-avanzada-en-baloncesto#:~:text=2PA%2C%203PA%3A%20Tiros%20de%20,OR%3A%20Rebotes%20ofensivos>

## Anexo

### Scripts de RStudio

```
ANÁLISIS DESCRIPTIVO.R x CORRELACIONES.R x CLUSTER.R x
Source on Save
1 library(dplyr)
2 library(e1071)
3 library(officer)
4 library(openxlsx)
5
6 ## Cargar la BBDD solo la T.21 y 22
7
8 nba_season_21_22 <- subset(BBDD_BASKET_NBA, TEMPORADA == "21-22")
9 equipos <- nba_season_21_22$Equipo (V)
10
11 nba_season_21_22_1 <- nba_season_21_22 %>%
12 select_if(is.numeric)
13 nba_season_21_22_analisis <- (nba_season_21_22_1 [, -1:-3])
14
15 # Calcular la media
16 mediaBBDD <- colMeans(nba_season_21_22_analisis)
17 # Calcular la desviación típica
18 desv_tipBBDD <- apply(nba_season_21_22_analisis, 2, sd)
19 # Calcular el mínimo
20 minimoBBDD <- apply(nba_season_21_22_analisis, 2, min)
21 # Calcular el máximo
22 maximoBBDD <- apply(nba_season_21_22_analisis, 2, max)
23 # Calcular el coeficiente de asimetría
24 coef_asimBBDD <- apply(nba_season_21_22_analisis, 2, skewness)
25 # Calcular el coeficiente de curtosis
26 coef_curtBBDD <- apply(nba_season_21_22_analisis, 2, kurtosis)
27
```

Figura 56: Script RStudio. Análisis Descriptivo.

```
27
28 ##### TABLA
29
30 mi_tabla <- data.frame(
31   Media = mediaBBDD,
32   DesviaciónTípica = desv_tipBBDD,
33   Mínimo = minimoBBDD,
34   Máximo = maximoBBDD,
35   CoeficienteAsimetría = coef_asimBBDD,
36   CoeficienteCurtosis = coef_curtBBDD
37 )
38
39 write.xlsx(mi_tabla, "mi_archivo.xlsx", sheetName = "Tabla1")
40
```

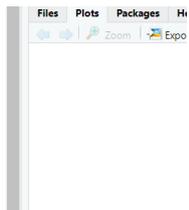


Figura 57: Script RStudio. Análisis Descriptivo (1).

```
57
58 ##TEST DE NORMALIDAD KOLMOGOROV-SMIRNOV
59 ks.test(nba_season_21_22_analisis_bueno, "pnorm", mediaBBDD, desv_tipBBDD)
60
```

Figura 58: Código RStudio Test Kolmogorov-Smirnov.

```
ANÁLISIS DESCRIPTIVO.R x MEDIA x CORRELACIONES.R x CLUSTER.R x
Source on Save
1 library(corrplot)
2 library(psych)
3
4 ##NORMALIZACIÓN DE VARIABLES
5 x <- nba_season_21_22_analisis
6 x_zscore <- scale(x)
7 x1 <- x_zscore
8
9 ### CORRELACIONES ENTRE VARIABLES
10 cor_matrix <- cor(x1)
11
12 ## GRÁFICO DE CORRELACIONES
13 corrplot(cor_matrix, method = 'number')
14 par(mar=c(0,0,0,0))
15 corrplot(cor_matrix, method = "color", type = "full", order = "hclust", tl.col = "black", tl.cex = 0.42, width = 50, height = 25)
16
17
18 # Guardar matriz como archivo csv
19 write.table(cor_matrix, file = "correlaciones_mi_BBDD.csv", sep = ",", quote = FALSE)
20
21
22
```

Figura 59: Script cálculo correlación.

```
21
22 # Análisis de componentes principales
23
24 resultado_pca <- princomp(x1)
25 summary(resultado_pca)
26 plot(resultado_pca)
27
28
```

**Figura 60:** Análisis PCA.

```
--
29 # Seleccionar los componentes principales que cumplen con el criterio de kaiser
30 componentes_kaiser <- resultado_pca$sdev^2 > 1
31
32 # Verificar los componentes seleccionados
33 componentes_kaiser
34
35 # Obtener los datos de los componentes principales seleccionados
36 datos_seleccionados <- resultado_pca$scores[, componentes_kaiser]
37
```

**Figura 61:** Criterio de Kaiser.

```
38 # Obtener los componentes principales utilizando el método de kaiser
39 pc <- principal(x1, nfactors = 14, rotate = "none")
40
41 # Realizar la rotación varimax
42 pc_rot <- principal(x1, nfactors = 14, rotate = "varimax")
43
44 # Obtener las cargas factoriales después de la rotación varimax
45
46 loadings <- loadings(pc_rot)
47
48 write.table(loadings, file = "loadings.txt", sep = "\t", row.names = TRUE)
49
50
```

**Figura 62:** Obtención cargas factoriales.

```
1 library(cluster)
2 library(factoextra)
3
4 # Convertir variables no numéricas a numéricas si es necesario
5 nba_season_21_22_analisis_bueno <- sapply(nba_season_21_22_analisis_bueno, as.numeric)
6
7 # Calcular la matriz de distancias
8 dist_matrix <- dist(nba_season_21_22_analisis_bueno)
9
10 # Generar el dendrograma
11 dend <- hclust(dist_matrix, method = "ward.D2")
12
13 # Graficar el dendrograma
14 plot(dend, main = "Dendrograma de clustering")
15
```

**Figura 63:** Método de Ward.

```
1 library(cluster)
2 library(factoextra)
3
4 # Convertir variables no numéricas a numéricas
5 nba_season_21_22_analisis_bueno <- sapply(nba_season_21_22_analisis_bueno, as.numeric)
6
7 # verificar si hay valores faltantes en los datos
8 missing_values <- complete.cases(nba_season_21_22_analisis_bueno)
9
10 # Eliminar las filas con valores faltantes (si es necesario)
11 nba_season_21_22_analisis_bueno <- nba_season_21_22_analisis_bueno[missing_values, ]
12
13 componentes <- nba_season_21_22_analisis_bueno
14
15 # Calcular la suma de cuadrados dentro del grupo para diferentes valores de k
16 k_max <- 10
17 ssw <- vector("double", k_max)
18 for (k in 1:k_max) {
19   kmeans_res <- kmeans(componentes, centers = k, nstart = 25)
20   ssw[k] <- kmeans_res$tot.withinss
21 }
22
23 # Graficar la curva del codo con ejes
24 plot(1:k_max, ssw, type = "b", pch = 19, frame = FALSE, xlab = "", ylab = "",
25      xlim = c(1, k_max), ylim = range(ssw))
26 axis(side = 1, at = 1:k_max, labels = 1:k_max)
27 axis(side = 2)
28 title(xlab = "Número de clústeres (k)", ylab = "Suma de cuadrados dentro del grupo (SSW)")
29
```

Figura 64: Método del codo.

```
31 # Encontrar el punto de codo
32 diff_ssw <- c(0, diff(ssw))
33 codo <- which(diff_ssw == max(diff_ssw)) + 1
34 points(codo, ssw[codo], col = "red", pch = 19)
35 text(codo, ssw[codo], labels = paste("codo:", codo), pos = 4)
36
37 # Ejecutar el análisis de k-means con los datos actualizados
38 componentes <- as.matrix(nba_season_21_22_analisis_bueno)
39 k_values <- c(2, 3, 4)
40 ssw <- vector("double", length(k_values))
41 assignments <- list()
42
43 # Calcular la suma de cuadrados dentro del grupo para diferentes valores de k
44 k_values <- c(2, 3, 4)
45 ssw <- vector("double", length(k_values))
46 assignments <- list()
47
48 for (i in 1:length(k_values)) {
49   k <- k_values[i]
50   kmeans_res <- kmeans(componentes, centers = k, nstart = 25)
51   ssw[i] <- kmeans_res$tot.withinss
52   assignments[[i]] <- kmeans_res$cluster
53 }
54
55 # Graficar la curva del codo con ejes
56 plot(k_values, ssw, type = "b", pch = 19, frame = FALSE, xlab = "Número de clústeres (k)", ylab = "Suma de cuadrados dentro del grupo (SSW)",
57      xlim = c(1, k_values), ylim = range(ssw))
58 points(k_values, ssw, col = "red", pch = 19)
59 text(k_values, ssw, labels = paste("codo:", k_values), pos = 4)
```

Figura 65: Método del codo (1).

```

60 # Imprimir la distribución de registros en cada clúster
61 for (i in 1:length(k_values)) {
62   k <- k_values[i]
63   asignaciones <- assignments[[i]]
64   distribucion <- table(asignaciones)
65   tabla_distribucion <- data.frame(Cluster = 1:length(distribucion), Registros = distribucion)
66   tabla_distribucion <- tabla_distribucion[order(tabla_distribucion$Cluster), ]
67   cat("Distribución de registros para", k, "clústeres:\n")
68   print(tabla_distribucion)
69   cat("\n")
70 }
71 }
72
73 componentes <- loadings[, c("RC1", "RC2", "RC3", "RC4")]
74
75 # Clusterización con 3 clústeres
76 k_2 <- 2
77 clusters_2 <- kmeans(componentes, centers = k_2, nstart = 25)
78
79 # Obtener los scores de cada clúster en cada componente principal
80 scores_2 <- aggregate(componentes, by = list(clusters_2$cluster), FUN = mean)
81 colnames(scores_2) <- c("cluster", "RC1", "RC2", "RC3", "RC4")
82
83 # Imprimir resultados
84 print(scores_2)

```

Figura 66: Método del codo (2).

```

1 # Obtener la variable de resultados (victoria/derrota) de los partidos
2 resultados <- BBDD_BASKET_NBA$`V/D`
3
4 # Obtener el número de observaciones en los datos
5 num_observaciones <- min(length(resultados), length(clusters_2$cluster))
6
7 # Crear un data frame con los resultados y los clústeres
8 df_resultados_clusters <- data.frame(Resultados = resultados[1:num_observaciones], Cluster = clusters_2$cluster[1:num_observaciones])
9
10 # Calcular la proporción de partidos ganados y no ganados por clúster
11 proporciones <- aggregate(Resultados ~ Cluster, data = df_resultados_clusters, FUN = function(x) {
12   ganados <- sum(x == "v")
13   perdidos <- sum(x == "d")
14   prop_ganados <- ganados / (ganados + perdidos)
15   prop_perdidos <- perdidos / (ganados + perdidos)
16   c(Ganados = prop_ganados, Perdidos = prop_perdidos)
17 })
18 proporciones$cluster <- factor(proporciones$cluster, levels = 1:k_2)
19
20 # Imprimir proporciones de partidos ganados y no ganados por clúster
21 cat("Proporción de partidos ganados y no ganados por clúster:\n")
22 print(proporciones)
23
24 # Realizar la prueba de chi-cuadrado
25 chi_test <- chisq.test(df_resultados_clusters$Resultados, df_resultados_clusters$cluster)
26 cat("\n")
27 cat("Prueba de Chi-cuadrado:\n")
28 print(chi_test)
29

```

Figura 67: Proporción V/D en cada clúster.

```

1 # Paso 1: Filtrar las estadísticas del equipo local
2 vic_der <- nba_season_21_22$`V/D`
3 columnas_a_eliminar_analisis_visitante <- c(1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 40, 41, 42, 43, 44, 45, 47, 49, 51, 53, 55, 57, 59,
4   61, 63, 65, 67, 69)
5 nba_season_21_22_analisis_bueno <- nba_season_21_22_analisis[, -columnas_a_eliminar_analisis_visitante]
6 # Agregar la columna "vic_der" a los datos "nba_season_21_22_analisis_bueno"
7 nba_season_21_22_analisis_bueno <- cbind(nba_season_21_22_analisis_bueno, vic_der)
8 nba_season_21_22_analisis_bueno$vic_der <- as.factor(nba_season_21_22_analisis_bueno$vic_der)
9
10
11
12 # Paso 2: Realizar análisis discriminante
13 library(MASS)
14 # Realizar el análisis discriminante
15 discrim_model <- lda(vic_der ~ ., data = nba_season_21_22_analisis_bueno)
16
17 # Calcular la matriz de confusión
18 library(caret)
19
20 # Convertir vic_der a factor
21 nba_season_21_22_analisis_bueno$vic_der <- factor(nba_season_21_22_analisis_bueno$vic_der, levels = c("D", "V"))
22
23 # Obtener las predicciones del modelo discriminante
24 predictions <- predict(discrim_model)
25
26 # Calcular la matriz de confusión
27 confusion <- confusionMatrix(data = as.factor(predictions$class), reference = nba_season_21_22_analisis_bueno$vic_der)
28
29 # Mostrar la matriz de confusión
30 print(confusion)
31
32
33 # Paso 4: Calcular las estadísticas de éxito en la predicción
34 success_rate <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
35 error_rate <- 1 - success_rate

```

Figura 68: Análisis discriminante.

```
33 # Paso 4: Calcular las estadísticas de éxito en la predicción
34 success_rate <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
35 error_rate <- 1 - success_rate
36
37 # Paso 5: Calcular los coeficientes lineales de las variables
38 linear_coefs <- discrim_model$scaling
39
40 # Paso 6: Calcular la curva ROC
41 library(pROC)
42
43 # Convertir vic_der a factor
44 nba_season_21_22_analisis_bueno$vic_der <- factor(nba_season_21_22_analisis_bueno$vic_der, levels = c("D", "V"))
45
46 # Ajustar el modelo discriminante
47 discrim_model <- lda(vic_der ~ ., data = nba_season_21_22_analisis_bueno)
48
49 # Obtener las probabilidades posteriores
50 posterior_probs <- as.data.frame(predict(discrim_model, nba_season_21_22_analisis_bueno)$posterior)
51
52 # Extraer las probabilidades para la clase "V"
53 posterior_V <- posterior_probs$V
54
55 # Calcular la curva ROC
56 roc_obj <- roc(response = nba_season_21_22_analisis_bueno$vic_der, predictor = posterior_V)
57
58 # Graficar la curva ROC
59 plot(roc_obj, main = "Curva ROC", xlab = "Tasa de Falsos Positivos", ylab = "Tasa de Verdaderos Positivos")
60
61 # Calcular el área bajo la curva ROC
62 auc_value <- auc(roc_obj)
63
64 # Imprimir el valor del AUC
65 print(auc_value)
```

Figura 69: Análisis discriminante (2)

```
68 # Paso 8: Graficar el resultado del modelo discriminante con el éxito del partido
69
70 # Obtener los valores de predicción del modelo discriminante
71 predicciones <- predict(discrim_model)$class
72
73 # Crear un dataframe con los resultados del modelo y el éxito del partido
74 data_grafica <- data.frame(Resultado = predicciones, Exito_Partido = nba_season_21_22_analisis_bueno$vic_der)
75
76 # Calcular la tabla de frecuencia de cada combinación de resultado y éxito del partido
77 tabla_grafica <- table(data_grafica$Resultado, data_grafica$Exito_Partido)
78
79 # Crear el gráfico de barras
80 barplot(tabla_grafica, beside = TRUE, legend = TRUE, col = c("red", "green"),
81         names.arg = c("Derrota", "Victoria"), xlab = "Resultado del modelo",
82         ylab = "Frecuencia", main = "Resultado del modelo vs. Exito del partido")
83
84 # Paso 9: Peso de las variables en las predicciones
85
86 # Paso 9: Peso de las variables en las predicciones
87
88 # Obtener los coeficientes lineales del modelo discriminante
89 coeficientes <- discrim_model$scaling
90
91 # Mostrar los pesos de las variables
92 print(coeficientes)
93
```

Figura 70: Análisis discriminante (3)



**ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA  
AGENDA 2030**

**Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los  
Objetivos de Desarrollo Sostenible de la agenda 2030.**

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

<b>Objetivos de Desarrollo Sostenibles</b>	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>	<b>No Procede</b>
ODS 1. <b>Fin de la pobreza.</b>				X
ODS 2. <b>Hambre cero.</b>				X
ODS 3. <b>Salud y bienestar.</b>				X
ODS 4. <b>Educación de calidad.</b>				X
ODS 5. <b>Igualdad de género.</b>				X
ODS 6. <b>Agua limpia y saneamiento.</b>				X
ODS 7. <b>Energía asequible y no contaminante.</b>				X
ODS 8. <b>Trabajo decente y crecimiento económico.</b>				X
ODS 9. <b>Industria, innovación e infraestructuras.</b>			X	
ODS 10. <b>Reducción de las desigualdades.</b>				X
ODS 11. <b>Ciudades y comunidades sostenibles.</b>				X
ODS 12. <b>Producción y consumo responsables.</b>				X
ODS 13. <b>Acción por el clima.</b>				X
ODS 14. <b>Vida submarina.</b>				X
ODS 15. <b>Vida de ecosistemas terrestres.</b>				X
ODS 16. <b>Paz, justicia e instituciones sólidas.</b>				X
ODS 17. <b>Alianzas para lograr objetivos.</b>			X	

Descripción de la alineación del TFG/TFM con los ODS con un grado de relación más alto.

\*\*\*Utilice tantas páginas como sea necesario.



**Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030.** (Numere la pàgina)

Tras ver las ODS y mi TFG, la alineación que presenta es bastante baja. Lo único destacable que he podido rescatar ha sido, que al final el mundo del baloncesto, en este caso la NBA, es una industria muy grande en la que evidentemente presenta este tipo de innovaciones, como el análisis de datos, que ayuda a las empresas, en este caso, los clubes a obtener mejores resultados.

En cuanto, a las alianzas para lograr objetivos, podemos alinearlos con el hecho de los contratos entre clubes, jugadores, staff técnico y todos los encargados para que un equipo funcione correctamente.

En definitiva, el trabajo no está muy alineado con las ODS de la agenda de 2030, ya que es un trabajo muy estadístico, donde no podemos evaluar estas competencias.