

# **Trabajo de Fin de Grado**

Grado en Ingeniería Informática

---

## **Propuesta de metodología para el análisis del rendimiento deportivo.**

### **Caso de uso baloncesto.**

*Proposal of methodology for the analysis of sports  
performance. Basketball use case*

Sergio Moreno Martín

---

La Laguna, 7 de *Septiembre* de 2021

D. **Dagoberto Castellanos Nieves**, con N.I.F. 79.234.766-L profesor Titular de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

**C E R T I F I C A ( N )**

Que la presente memoria titulada:

*“Propuesta de metodología para el análisis del rendimiento deportivo. Caso de uso baloncesto”*

Ha sido realizada bajo su dirección por D. **Sergio Moreno Martín**, con N.I.F. 54.109.477-E.

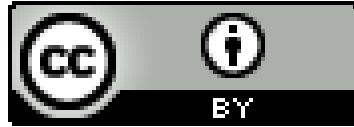
Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 7 de septiembre de 2021.

## **Agradecimientos**

En primer lugar, quiero agradecer a mi tutor, Dagoberto Castellano Nieves, por guiarme y aconsejarme en la elaboración de este trabajo, por tener la paciencia durante las asignaturas que me ha impartido clases y principalmente con este trabajo final de grado.

A mis amigos por estar siempre ahí en los momentos difíciles y por último a mi familia, puesto que sin ellos nunca habría llegado a donde estoy ahora mismo.

## Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

## Resumen

*El objetivo principal de este trabajo ha sido plantear una metodología para el análisis del rendimiento deportivo centrado en el baloncesto.*

*Utilizando estadísticas sencillas y avanzadas del baloncesto, se propone definir una metodología que permita a los jugadores, entrenadores y otros integrantes del staff técnico de los equipos, mediante el uso de la ciencia de datos y la inteligencia artificial, analizar y mejorar el rendimiento durante los entrenamientos y partidos, simultáneamente agilizar la toma de decisiones en ciertos momentos claves de los partidos. Todo este proceso tiene como objetivo final la creación de gráficos que permitan a cualquier persona interpretar los resultados de forma correcta.*

**Palabras clave:** baloncesto, ciencia de datos, inteligencia artificial, estadísticas, gráficos, rendimiento

## **Abstract**

*The main objective of this work has been to propose a methodology for the analysis of sports performance focused on basketball.*

*Using simple and advanced basketball statistics, it is proposed to define a methodology that allows players, coaches and other members of the teams' technical staff, through the use of data science and artificial intelligence, to analyze and improve performance during the games. training sessions and matches, simultaneously streamline decision-making at certain key moments in matches. The final objective of all this process is to create graphs that allow anyone to interpret the results correctly.*

**Keywords:** *basketball, data science, artificial intelligence, statistics, graphs, performance.*

# ÍNDICE GENERAL

<b>ÍNDICE DE FIGURAS</b>	<b>8</b>
<b>ÍNDICE DE TABLAS</b>	<b>9</b>
<b>INTRODUCCIÓN</b>	<b>10</b>
<b>ESTADO DEL ARTE</b>	<b>12</b>
ANTECEDENTES	12
TECNOLOGÍAS EMPLEADAS	12
Inteligencia artificial	12
Ciencia de datos	13
LENGUAJE DE PROGRAMACIÓN	14
HERRAMIENTAS Y LIBRERÍAS	14
ORIGEN DE LOS DATOS	14
NBA	15
Grupo Constructalia Tacoronte y Cajasiete Canarias	16
DEFINICIÓN DE LA METODOLOGÍA	17
<b>OBJETIVOS</b>	<b>19</b>
OBJETIVO GENERAL DEL PROYECTO	19
OBJETIVOS ESPECÍFICOS	19
Análisis de las estadísticas	19
Visualización y generación del informe	19
Aplicación en los jugadores	19
<b>DESARROLLO</b>	<b>20</b>
CONFIGURACIÓN DEL ENTORNO	20
ENTORNO DE DESARROLLO	20
PROCESAMIENTO DE LOS DATOS	20
IMPLEMENTACIÓN DE LA METODOLOGÍA	20
<b>CONCLUSIONES Y LÍNEAS FUTURAS</b>	<b>28</b>
CONCLUSIONES	28
LÍNEAS FUTURAS	28
<b>SUMMARY AND CONCLUSIONS</b>	<b>29</b>
CONCLUSIONS	29
FUTURE LINES	29
<b>PRESUPUESTO</b>	<b>30</b>
PRESUPUESTO DE RECURSOS HUMANOS	30
PRESUPUESTO MATERIAL	30
PRESUPUESTO FINAL	31
<b>APÉNDICES</b>	<b>32</b>
SCRIPT PARA EXTRAER LOS PTS ANOTADOS EN CADA ZONA DE LA	

CANCHA Y REPRESENTARLOS GRÁFICAMENTE	32
ANÁLISIS CON SWEETVIZ	33
ANÁLISIS INDIVIDUAL T1% - PTS	33
<b>REFERENCIAS</b>	<b>35</b>



## ÍNDICE DE FIGURAS

Figura 1. Estadísticas Catch & Shoot de nba.stats.com	16
Figura 2. Estadística del Grupo Constructalia Tacoronte	17
Figura 3. Estadísticas del Cajasiete Canarias	17
Figura 4. Flujo de trabajo de la metodología	18
Figura 5. Mapa de calor PTS de Luka Doncic y Damian Lillard Regular Season	21
Figura 6. Mapa de calor PTS Stephen Curry Regular Season	21
Figura 7. Mapa de calor Luka Doncic Regular Season vs Clutch 30s Playoff	22
Figura 8. Mapa de calor Damian Lillard Regular Season vs Clutch 30s Playoff	23
Figura 9. Análisis de PTS entre Grupo Constructalia Tacoronte y RCNT	24
Figura 10. Análisis de T3I entre Grupo Constructalia Tacoronte y RCNT	25
Figura 11. Análisis de T3A entre Grupo Constructalia Tacoronte y RCNT	25
Figura 12. Relación entre T1% y PTS para los jugadores Diego, Angel y David del Cajasiete Canarias	27

## **ÍNDICE DE TABLAS**

Tabla 1. Presupuesto de recursos Humanos	30
Tabla 2. Presupuesto de material	30
Tabla 3. Presupuesto final	31

## Capítulo 1 INTRODUCCIÓN

En un primer momento destaca la importancia del análisis de los jugadores centrado en los aspectos antropométricos, perfil fisiológico y aptitudes de los jugadores.

El rendimiento en deportes como el fútbol o el baloncesto va más allá del estudio de las características de los deportistas, debido a que, el rendimiento deportivo no lo materializan los equipos que cuentan con los mejores jugadores.

La lógica de los deportes de equipo, la presencia de varios jugadores, etc. obliga a realizar estudios que se centren en el conocimiento del propio deporte, al igual que lo ocurrido durante la propia acción del juego. Desde el estudio de lo sucedido en la competición y el entrenamiento se empieza a descubrir algunas de las claves del rendimiento de equipo.

Durante décadas, los entrenadores de los deportes colectivos han ido registrando las acciones de juego que consideraban más relevantes para su posterior análisis. Con el transcurso del tiempo, los registros comenzaron a normalizarse, estandarizarse y universalizarse, permitiendo a los entrenadores, investigadores y directores deportivos estudiar con profundidad el desarrollo del propio juego.

Cuando observamos partidos de baloncesto de alto nivel podemos apreciar que a los entrenadores, en situaciones de tiempo muerto, final de cuarto, o detención del partido, se les entrega unos documentos en los que se recogen las estadísticas del partido. Los entrenadores observan y tratan de encontrar algún dato clave para realizar los cambios oportunos en su sistema de juego, o poder transmitir a sus jugadores las pautas a seguir para conseguir la victoria.

Visto desde una perspectiva sin mucho conocimiento sobre el baloncesto, puede parecer que los valores en cada una de las estadísticas no aportan gran información, pero en su conjunto ayudan a comprender como se ha rendido en una competición, así como establecen unas pautas para poder mejorar en los entrenamientos.

La investigación fue avanzando tratando de comparar el rendimiento de los jugadores y equipos de diferentes ligas o categorías, usando indicadores de partido, de modo que se pudiera encontrar los factores que definen los equipos o a los jugadores. En este contexto de análisis, el factor diferencial es el estudio de la competición, puesto que se considera el mejor momento para evaluar el rendimiento en situación real de partido. Dicho estudio permite identificar a los mejores de los peores (equipos y/o jugadores), y consecuentemente alcanzar mejores resultados deportivos.

Uno de los aspectos importantes en el análisis del rendimiento deportivo es la evolución del reglamento. Durante la historia del baloncesto, el reglamento ha evolucionado en multitud de ocasiones, por ejemplo la aparición de la línea de 3 puntos en el baloncesto americano en el año 1945, con su posterior incorporación por parte de la FIBA (Federación Internacional de

Baloncesto) en el año 1984, tras la finalización de los juegos olímpicos de Los Ángeles. La distancia de la línea de triple se vio incrementada de 6,25 cm a 6,75 cm en el año 2010 en el baloncesto FIBA, provocando que durante los primeros años de su aplicación se realizaron estudios para medir las capacidades anotadoras desde el triple. Dichos análisis reflejaron una disminución de los porcentajes en jugadores que no son especialistas en el tiro. Actualmente el baloncesto ha evolucionado hacia un estilo de juego más rápido y de mayor anotación, esto significa que los jugadores ponen más esfuerzos en mejorar sus tiros de larga distancia.

Por otro lado, la estadística oficial supone una de las herramientas más potentes que se puede utilizar como medio de control para la competición y desarrollo de los jugadores.

Hace unas décadas, las estadísticas eran precarias, insuficientes y con gran dificultad de recopilación, por suerte la evolución de la tecnología ha facilitado la recopilación de los datos. La NBA (National Basketball Association) es la mayor liga de baloncesto del mundo, pionera en la recopilación de datos, utilizan cámaras situadas por toda la cancha y con la ayuda de **Inteligencia Artificial (IA)** y **Machine Learning (ML)** obtienen datos de movimiento, toques de balón, localización de los jugadores en la misma, etc. El nivel de recopilación y análisis de los datos en la NBA ha llegado al punto de monitorizar, con sensores de proximidad, la situación de la pandemia mundial (COVID19) entre sus equipos, no solo durante los partidos, sino también durante viajes, entrenamientos, tiempo en las instalaciones o actos de caridad. .

Los equipos de nivel profesional han realizado durante los últimos 15 años una gran inversión en el análisis de los datos con el fin de mejorar el rendimiento de los jugadores, prevenir lesiones, estudiar movimientos de mercado, etc. A día de hoy es muy poco común ver un cuerpo técnico de un club que no cuente con científicos de datos, analistas, ojeadores, expertos en **Big Data**, trabajando lado a lado con los entrenadores y directores deportivos.

## Capítulo 2 ESTADO DEL ARTE

### 2.1 ANTECEDENTES

Hoy en día los científicos de datos utilizan técnicas de **inteligencia artificial [1]** y **análisis predictivo [2]** para ahorrar tiempo y recursos, especialmente a la hora de interactuar con grandes volúmenes de datos (estáticos o dinámicos), favoreciendo a una mejora en la toma de decisiones.

Inicialmente su uso se limitaba al mundo de las apuestas y las retransmisiones televisivas, pero actualmente es una realidad en los equipos de élite de cualquier deporte. En nuestro caso en concreto, el **análisis de datos** en el baloncesto ha experimentado un rápido crecimiento en los últimos 15 años. El desarrollo de **herramientas y técnicas predictivas [3]** comenzó a medir mejor el rendimiento tanto del jugador como del equipo.

Una de las ligas referente en la recopilación y uso de los datos es la **NBA**, donde un elevado número de toma de decisiones, tanto administrativas y de gestión como deportivas se basan en el análisis. Los equipos utilizan el análisis principalmente para el diseño de estrategias ganadoras, predecir/evitar lesiones y realizar una exploración del mercado.

El acceso a herramientas estadísticas centradas en un deporte como el baloncesto, que permitan un posterior análisis de los datos de una forma eficiente y eficaz, ha mejorado con el paso del tiempo, dejando de lado su uso únicamente profesional logrando de esta forma acercar al deporte amateur a un nivel de estudio impensable años atrás.

### 2.2 TECNOLOGÍAS EMPLEADAS

#### 2.2.1 *Inteligencia artificial*

Con el rápido desarrollo de la informática, la tecnología de la **inteligencia artificial** crece a pasos agigantados y su aplicación en competiciones deportivas se ha vuelto cada vez más generalizada. Mediante **aprendizaje automático [4]**, se han conseguido grandes avances en la evolución del juego, esto ha permitido a entrenadores predecir y cuantificar con precisión el rendimiento individual o de equipo. Se trata de la tecnología perfecta para el análisis de situaciones reales de partido, ya que es apropiada para identificar tendencias.

A través de esto, el análisis de innumerables situaciones de partidos, como **pick-and-rolls** y **aclarados**, se vuelven más fáciles. A medida que se recopilan más datos, el aprendizaje automático puede identificar variantes nunca antes vistas, lo que hace que el **análisis en tiempo real [5]** sea eficiente.

Algunas de las herramientas de aprendizaje automático que son útiles en el baloncesto en la actualidad y en el futuro son el procesamiento del lenguaje natural y las series temporales.

Las series de tiempo tienen en cuenta datos cronológicos. El objetivo es encontrar estadísticas, patrones, conocimientos y características entre los puntos de datos para producir

predicciones de datos. En esencia, la herramienta proporciona un potencial ilimitado para identificar las fortalezas y debilidades de cada jugador en un equipo.

La inteligencia artificial ha revolucionado las **transmisiones en vivo** y las **apuestas deportivas**. Para la transmisión, una herramienta que recopila datos automáticamente durante los juegos, sin participación humana, puede revolucionar la forma en que se realiza la transmisión en vivo e influir en la forma en que los fanáticos ven el deporte. La IA hace posible que los canales publiquen información en vivo sobre el rendimiento de los equipos o jugadores, o incluso muestran estadísticas importantes que son difíciles de detectar por los humanos en tiempo real.

Al utilizar la historización impulsada por IA, es posible comparar las coincidencias y los rendimientos actuales con los anteriores. Esto mejora la experiencia de los clientes frente a su canal de televisión y aumenta su compromiso.

Otra industria que aprovecha la IA es el sector de las apuestas deportivas. Sin embargo, las casas de apuestas tienen diferentes formas de utilizarla para aumentar la eficiencia o para ganar más clientes.

Ya está marcando el comienzo del baloncesto en el futuro. La tecnología permite un mejor y más rápido desarrollo de los jugadores y proporciona a los entrenadores de baloncesto información precisa. Los entrenadores pueden pasar esta información a los miembros de su equipo.

Esto es crucial porque no solo promueve un aprendizaje más rápido, sino que también simplifica la recopilación de información útil que se utilizará en el desarrollo de los jugadores.

### 2.2.2 *Ciencia de datos*

La **ciencia de datos** [6] es otro punto fundamental en la actualidad deportiva y combina múltiples campos que incluyen **estadística, métodos científicos, IA y análisis de datos** para extraer valor de los datos. Abarca la preparación de datos para su análisis, incluidas las tareas de **limpieza, agregación y manipulación** de los datos para realizar análisis avanzados de los datos.

El proceso de analizar y utilizar los datos es iterativo en lugar de lineal, pero este es el modo en que el ciclo de vida fluye normalmente en un proyecto de modelado de datos:

Planificación, construcción de los modelos de datos, evaluación de los modelos, explicación de los modelos, implementación de un modelo, supervisión de los modelos.

## 2.3 LENGUAJE DE PROGRAMACIÓN

Se ha hecho uso del lenguaje de programación **Python** [7]. A día de hoy es uno de los lenguajes más utilizados y lidera los rankings de popularidad, gracias a su facilidad de uso, versatilidad y curva de aprendizaje baja.

## 2.4 HERRAMIENTAS Y LIBRERÍAS

La herramienta elegida para el proyecto ha sido **Jupyter Notebook** [8], una aplicación web de código abierto que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Los usos incluyen: limpieza y transformación de datos, simulación numérica, modelado estadístico, visualización de datos, aprendizaje automático y mucho más.

**NBA API** [9]. Se trata de un Cliente API para acceder a las APIs de [www.nba.com](http://www.nba.com), permitiendo de esta forma tener acceso a toda la información disponible de una manera simplificada.

**Matplotlib** [10]. Biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python y su extensión matemática NumPy

**Pandas** [11]. Biblioteca de software escrita como extensión de *NumPy* para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales.

**Sweetviz** [12]. Biblioteca de Python de código abierto que genera visualizaciones de alta densidad para iniciar EDA (Análisis de datos exploratorios) con solo dos líneas de código. La salida es una aplicación HTML totalmente autónoma.

El sistema se basa en la visualización rápida de valores objetivo y la comparación de conjuntos de datos. Su objetivo es ayudar a realizar un análisis rápido de las características del objetivo, datos de entrenamiento frente a pruebas y otras tareas de caracterización de datos similares.

## 2.5 ORIGEN DE LOS DATOS

Los datos utilizados en el proyecto son estadísticas sencillas y avanzadas de partidos reales de baloncesto, tanto de nivel profesional como semiprofesional y de categoría base. Los datos con los que se trabaja son de dos tipos:

**Strings de caracteres:** representan los nombres y apellidos de los jugadores, nombres de los equipos, resultados, conferencias, etc.

**Numérico:** abarca desde las estadísticas más sencillas como los minutos jugados (MIN), puntos anotados (PTS), valoración (VAL), robos (STL) asistencias (AST), etc.

Se utilizan estadísticas simples como las mencionadas para calcular las conocidas como **estadísticas avanzadas**. Dentro de las estadísticas avanzadas se pueden destacar los siguientes valores:

Ritmo o "Pace" : representa el ritmo de juego del equipo, expresando la cantidad de posesiones por juego que utiliza.

$$\text{Ritmo} = \text{puntos} / \text{posesiones} * 100$$

Las posesiones son calculadas de la siguiente forma:

$$\text{Pos} = \text{FGA} - \text{OR} + \text{TO} + (\text{FTA} * 0.4)$$

*FGA: Lanzamientos de campo*

*OR: Rebotes ofensivos*

*TO: Pérdidas*

*FTA: Tiros libres lanzados*

Eficiencia ofensiva (eff of): Puntos por cada 100 posesiones.

$$\text{EFF OF} = (\text{puntos/posesiones}) * 100$$

Eficiencia defensiva (eff def): puntos recibidos cada 100 posesiones.

$$\text{EFF DEF} = (\text{puntos recibidos}/100 \text{ posesiones del rival}) * 100$$

Porcentaje efectivo de tiro de campo (eFG%): calcula el porcentaje de tiros de campo dándole un valor extra a los triples..

$$\text{eFG\%} = (\text{FGM} + 0.5 * 3PM) / \text{FGA}$$

### 2.5.1 NBA

La principal fuente de datos del proyecto son las estadísticas oficiales de la **NBA**. Extraídas de los partidos mediante un sistema de cámaras que hacen un seguimiento completo de todos los jugadores que se encuentran en la cancha durante un partido, accesible mediante su propia página web.



**Players Tracking Catch & Shoot** ▾

SEASON	SEASON TYPE	PER MODE	SEASON SEGMENT
2020-21	Regular Season	Per Game	All Games

Player Tracking is not available for all games, please see the [Tracking Status](#) page for a game's availability.

PLAYER	TEAM	GP	MIN	PTS	FGM	FGA	FG%
Aaron Gordon	DEN	50	27.7	2.7	1.0	2.5	39.0
Aaron Holiday	IND	66	17.8	2.4	0.8	2.1	38.7
Aaron Nesmith	BOS	46	14.5	2.4	0.8	2.0	39.8
Abdel Nader	PHX	24	14.8	2.3	0.8	1.8	42.9
Adam Mokoka	CHI	14	4.0	0.2	0.1	0.6	12.5
Al Horford	BOS	28	27.9	5.8	2.0	5.6	36.1
Al-Farouq Aminu	CHI	23	18.9	1.1	0.4	1.5	25.7
Alec Burks	NYK	49	25.6	3.4	1.1	3.0	38.1
Aleksej Pokusevski	OKC	45	24.2	3.2	1.1	3.7	28.7
Alen Smilagic	GSW	15	5.6	0.6	0.2	0.5	37.5
Alex Caruso	LAL	58	21.0	2.6	0.9	2.2	40.8

Figura 1. Estadísticas Catch & Shoot de nba.stats.com

Para la realización de este trabajo, se van a usar las estadísticas de puntos anotados en cada posición del campo durante la pasada temporada regular, al igual que los puntos anotados en situaciones “Clutch” (momento del partido en el que quedan menos de 5 minutos para finalizar y la diferencia entre equipos es de cinco puntos o menos), centrándonos únicamente en partidos de Playoffs con un tiempo restante de 30 segundos o menos.

### 2.5.2 Grupo Constructalia Tacoronte y Cajasiete Canarias

Se cuenta con datos de dos equipos pertenecientes a clubes de la isla

**Grupo Constructalia Tacoronte.** Equipo senior que juega en la Primera División Nacional Masculina, terceros de Canarias en la pasada temporada.

**Cajasiete Canarias.** Equipo de formación de categoría Infantil Masculina, octavo clasificado de España en la pasada temporada.



3. Análisis de los datos para evaluar el rendimiento.

Según la necesidad de cada situación se centrarán en análisis individuales, grupales.

4. Visualización de los resultados mediante herramientas gráficas.

Buscando la simplicidad a la hora de las visualizaciones facilitando su lectura en momentos críticos.

5. Evaluación de los resultados por parte de los expertos.

Se consideran expertos a los entrenadores, analistas de datos, científicos de datos, directores deportivos. Personas que puedan tomar decisiones en base a los resultados obtenidos.

6. Aplicación individual/grupal de los resultados.

Desarrollar entrenamientos con ejercicios específicos donde se trabajen las actitudes analizadas.

7. Comparación con los datos de origen.

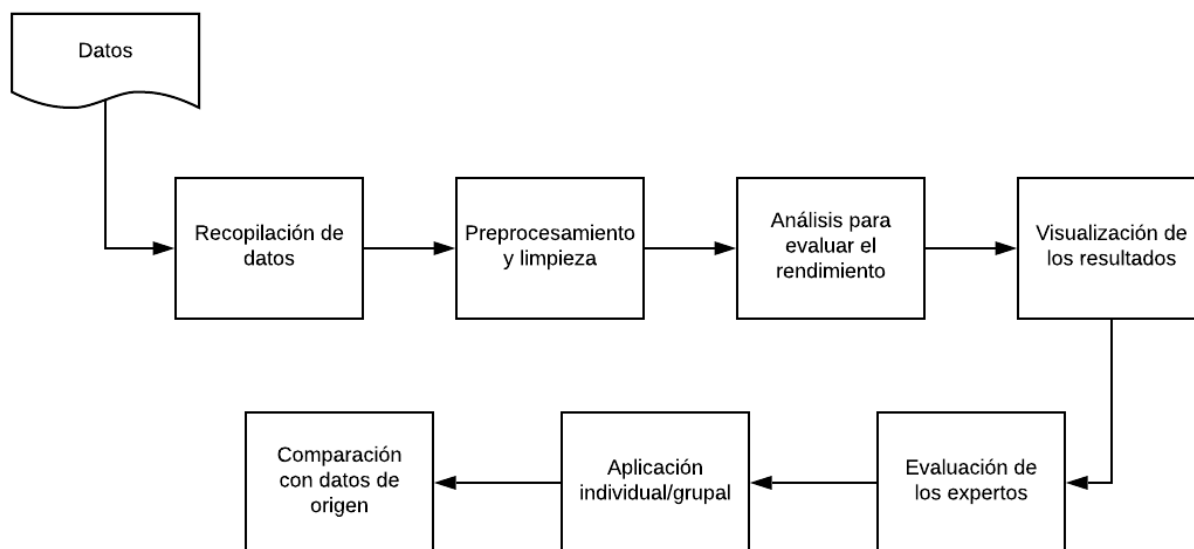


Figura 4 . Flujo de trabajo de la metodología

## **Capítulo 3 OBJETIVOS**

### **3.1 OBJETIVO GENERAL DEL PROYECTO**

El objetivo de este TFG es diseñar una metodología que guíe al usuario a obtener datos, procesarlos, analizarlos y finalmente mostrar los resultados obtenidos para su posterior aplicación sobre los jugadores durante los entrenamientos y partidos, independientemente del nivel al que se esté practicando el baloncesto, con el fin de mejorar el rendimiento individual y colectivo dentro de la cancha.

El propósito es acercar el análisis de rendimiento que tienen los equipos profesionales a equipos amateur o de formación, favoreciendo el desarrollo de sus aptitudes dentro del baloncesto. Se trata de un proceso completamente dependiente de los datos que se recopilan, los resultados serán visibles según la capacidad de aprendizaje y sacrificio de de los jugadores/equipos.

### **3.2 OBJETIVOS ESPECÍFICOS**

#### *3.2.1 Análisis de las estadísticas*

La finalidad es analizar las estadísticas obtenidas de las diferentes fuentes permitiendo sacar conclusiones de los puntos fuertes y débiles de los jugadores/equipos durante la práctica del baloncesto.

#### *3.2.2 Visualización y generación del informe*

Se propone generar gráficos que ayuden a visualizar de una forma óptima los resultados del análisis, permitiendo la fácil comprensión por parte de los implicados.

#### *3.2.3 Aplicación en los jugadores*

Se pretende que los resultados obtenidos tras el análisis de las estadísticas y la posterior visualización , se apliquen sobre los jugadores durante cada entrenamiento y partido, con el fin de mejorar día a día en los aspectos deseados. En este punto, queda en mano de los jugadores y entrenadores el aumento del rendimiento mediante ejercicios específicos como puede ser el pase, tiro, defensa, etc .

## Capítulo 4 DESARROLLO

### 4.1 CONFIGURACIÓN DEL ENTORNO

Para la realización del proyecto no es necesario definir un entorno de trabajo específico. Por comodidad y familiaridad, el desarrollo se ha realizado en un dispositivo con **Windows 10** corriendo **WSL2**.

### 4.2 ENTORNO DE DESARROLLO

Se ha instalado el editor de código **Visual Studio Code** y los plugins oficiales de Microsoft para Python y Jupyter Notebook.

### 4.3 PROCESAMIENTO DE LOS DATOS

Con las estadísticas obtenidas de la **NBA**, el procesamiento se ha realizado sobre el formato json resultante de las llamadas a la API **nba\_api** y posterior uso como un **DataFrame** de **Pandas**.

Las estadísticas del **Grupo Constructalia Tacoronte** y **Cajasiete Canarias** están en formato .xlsx. Para facilitar el trabajo, se le realiza una conversión a .csv eliminando las filas y columnas que pueden generar conflictos durante su lectura.

Se han transformado los valores de porcentajes a tipos numéricos para evitar problemas con su procesamiento.

### 4.4 IMPLEMENTACIÓN DE LA METODOLOGÍA

#### ***Caso 1: Estadísticas oficiales de la NBA Temporada Regular***

#### 1. Recopilación de los datos estadísticos

La recopilación de los datos se ha ejecutado mediante llamadas de la API **nba\_api** a la web de estadísticas de la NBA.

Se han seleccionado a 3 jugadores de la liga como **Luka Doncic**, **Damian Lillard** y **Stephen Curry**, se estudiará los puntos anotados en cada zona de la cancha durante la temporada regular y una comparación con los puntos anotados en los momentos decisivos del partido, con un tiempo inferior a 30 segundos.

#### 2. Preprocesamiento y limpieza del conjunto de datos

Los datos al ser de la fuente oficial de NBA están bien estructurados, por tanto el único preprocesamiento que hay que realizar sobre ellos es añadirlos a un DataFrame de Pandas.

#### 3. Análisis de los datos para evaluar el rendimiento

El análisis de los datos elegidos se realizará en el siguiente punto dado que nos centramos en datos con coordenadas de la cancha de baloncesto

#### 4. Visualización de los resultados

Para empezar, se crean los bordes para la simulación de la cancha de baloncesto, al tratarse de PTS durante la temporada regular y playoff, buscando situaciones reales en campo de ataque, evitamos la visualización del campo defensivo.

Posteriormente, se cargan los datos correspondientes de cada jugador.

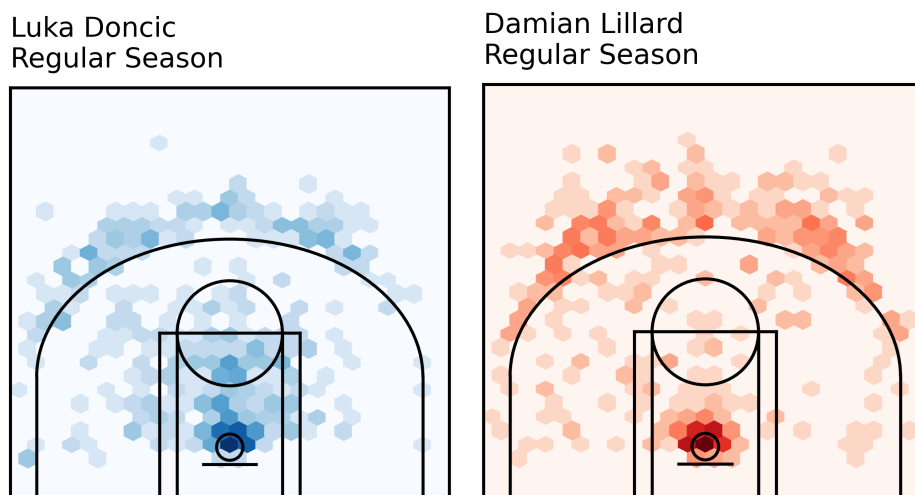


Figura 5. Mapa de calor PTS de Luka Doncic y Damian Lillard Regular Season

Curry  
Regular Season

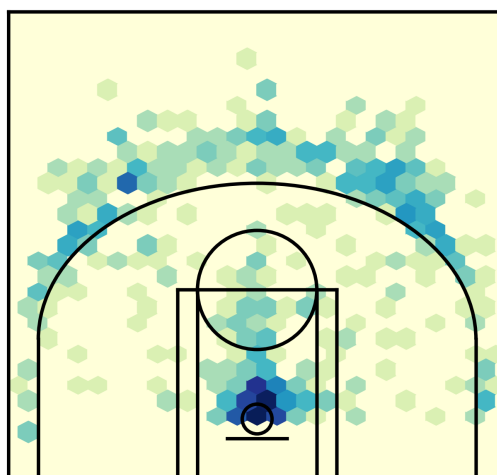


Figura 6. Mapa de calor PTS Stephen Curry Regular Season

#### 5. Evaluación de los resultados por parte de los expertos

Como entrenador de baloncesto de nivel I y estudiando los resultados del punto anterior, se observa claramente que el juego de los 3 jugadores estudiados es muy similar, utilizando una elevada cantidad de volumen de tiros desde la línea de 3 puntos. En el caso de Luka Doncic, hay un incremento de PTS desde el interior de la zona gracias a la diferencia de altura respecto a los otros dos jugadores. Un punto a mejorar es la anotación desde media distancia (4-5 metros) y los

lanzamientos desde las esquinas.

Con Damian Lillard se observa un patrón similar, con un aumento en los PTS de lanzamientos de larga distancia. Un punto a mejorar son los lanzamientos desde una distancia de 3-4 metros y los de media distancia desde el lado derecho de la cancha.

Con Stephen Curry se observa el mismo estilo de gráfico que los otros dos jugadores, añadiendo un mayor volumen de PTS desde las esquinas, debido a su juego sin balón. Un punto a mejorar sería la anotación desde la media distancia en la esquina izquierda de la cancha.

#### 6. Aplicación individual/grupal

Se planificaría ejercicios de tiro y finalizaciones desde las zonas descritas anteriormente durante las sesiones de entrenamiento.

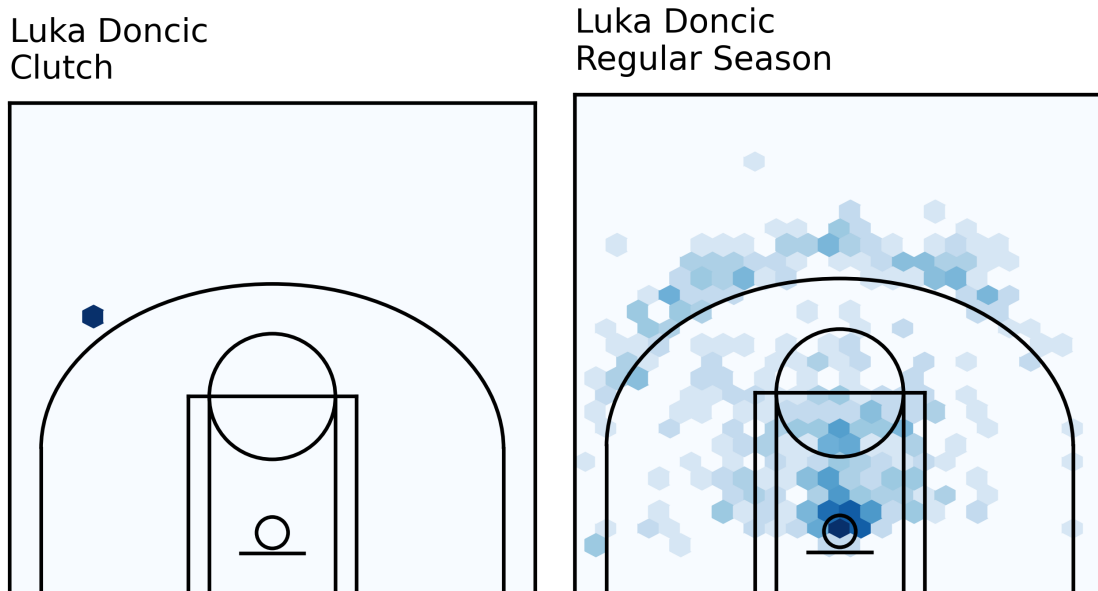
#### 7. Comparación con los datos de origen

Para finalizar se comparan los datos originales con los datos obtenidos después de realizar las sesiones de entrenamientos permitiendo ver si se ha realizado alguna mejora en su rendimiento.

### ***Caso 2: Estadísticas oficiales de la NBA Temporada Regular y Clutch Playoff***

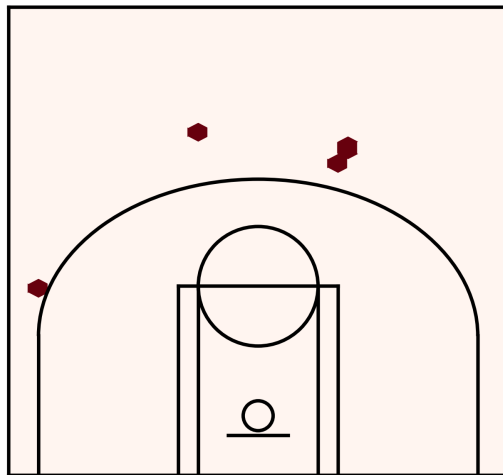
Al tratarse de un caso con el mismo origen de datos que el Caso 1, hasta el punto 4 se aplicarán de la misma forma.

Visualización de los resultados

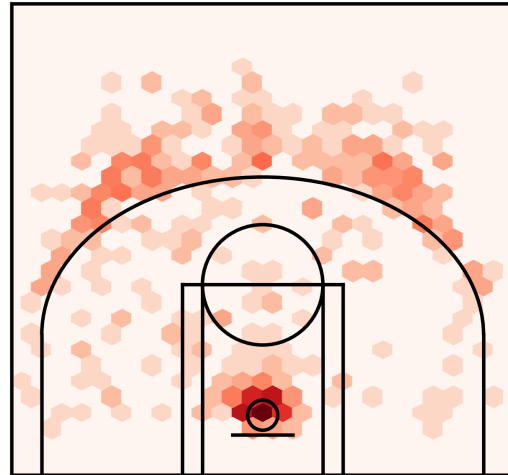


*Figura 7. Mapa de calor Luka Doncic Regular Season vs Clutch 30s Playoff*

Damian Lillard  
Clutch 30s Playoffs



Damian Lillard  
Regular Season



*Figura 8. Mapa de calor Damian Lillard Regular Season vs Clutch 30s Playoff*

### Evaluación de los resultados por parte de los expertos

Las gráficas reflejan que tanto Luka Doncic como Damian Lillard, anotan más desde las posiciones en las que se encuentran más cómodos en función a su juego.

### Aplicación individual/grupal

Realizar ejercicios durante los entrenamientos en los que se practiquen situaciones con tiempo y el marcador ajustado, variando la posición de los lanzamientos dependiendo de la defensa.

### Comparación con los datos de origen

Se realizará una comprobación de los PTS anotados antes y después de la aplicación de la metodología para ver su eficacia.

### **Caso 3: Estadísticas de equipos locales**

#### 1. Recopilación de los datos estadísticos

Para este caso, los datos han sido proporcionados por un entrenador de ambos clubes, contamos con las estadísticas de todos los partidos de la temporada.

#### 2. Preprocesamiento y limpieza del conjunto de datos

Los datos se han transformado desde .xlsx a .csv. Se han eliminado filas y columnas innecesarias y se han realizado modificaciones en su conjunto añadiendo valores que estaban vacíos y cambiando el tipo de algunos datos.

#### 3. Análisis de los datos para evaluar el rendimiento

En este caso nos centramos en las estadísticas simples como son minutos (MIN), puntos (PTS), Triples intentados (T3I) y Triples anotados (T3A).



#### 4. Visualización de los resultados

Para la visualización se ha utilizado la herramienta SweetViz que facilita la representación gráfica de los conjuntos de datos utilizados. Crea un informe visual detallado de las estadísticas de un partido.

Los datos representados corresponden a la comparación de los partidos que enfrentaron al Grupo Constructalia Tacoronte y RCNT. En color azul se representa el primer enfrentamiento entre ambos equipos, correspondiente a la sexta jornada disputada el 13 de Febrero de 2021, y en color naranja el segundo enfrentamiento, correspondiente a la octava jornada disputada el 27 de Febrero de 2021.

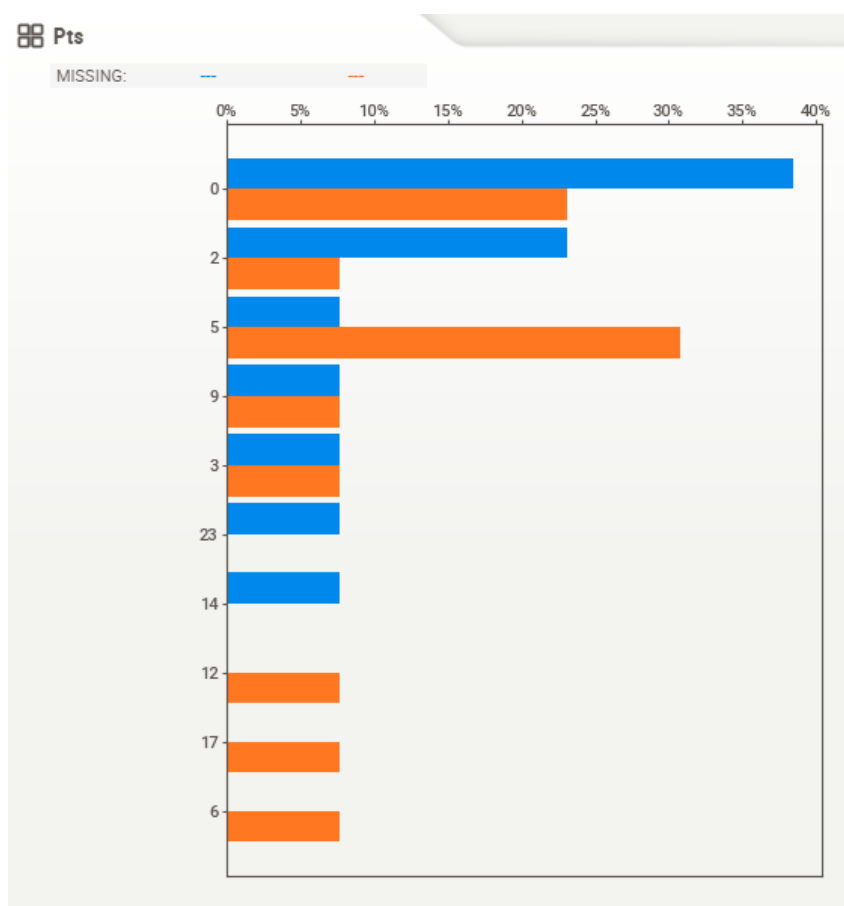


Figura 9. Análisis de PTS entre partidos del Grupo Constructalia Tacoronte y RCNT

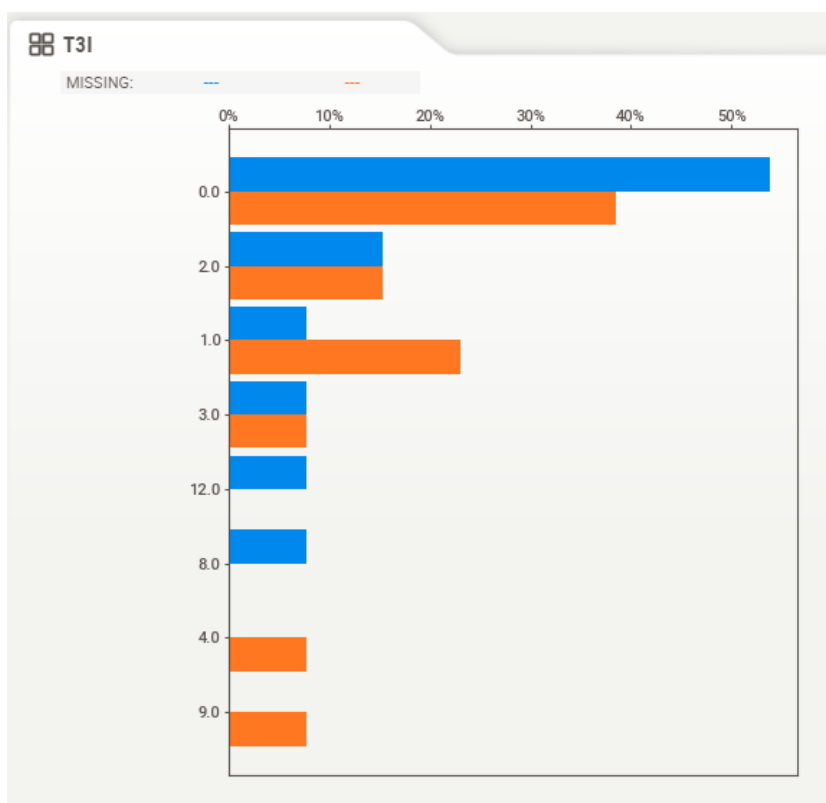


Figura 10. Análisis de T3I entre partidos del Grupo Constructalia Tacoronte y RCNT

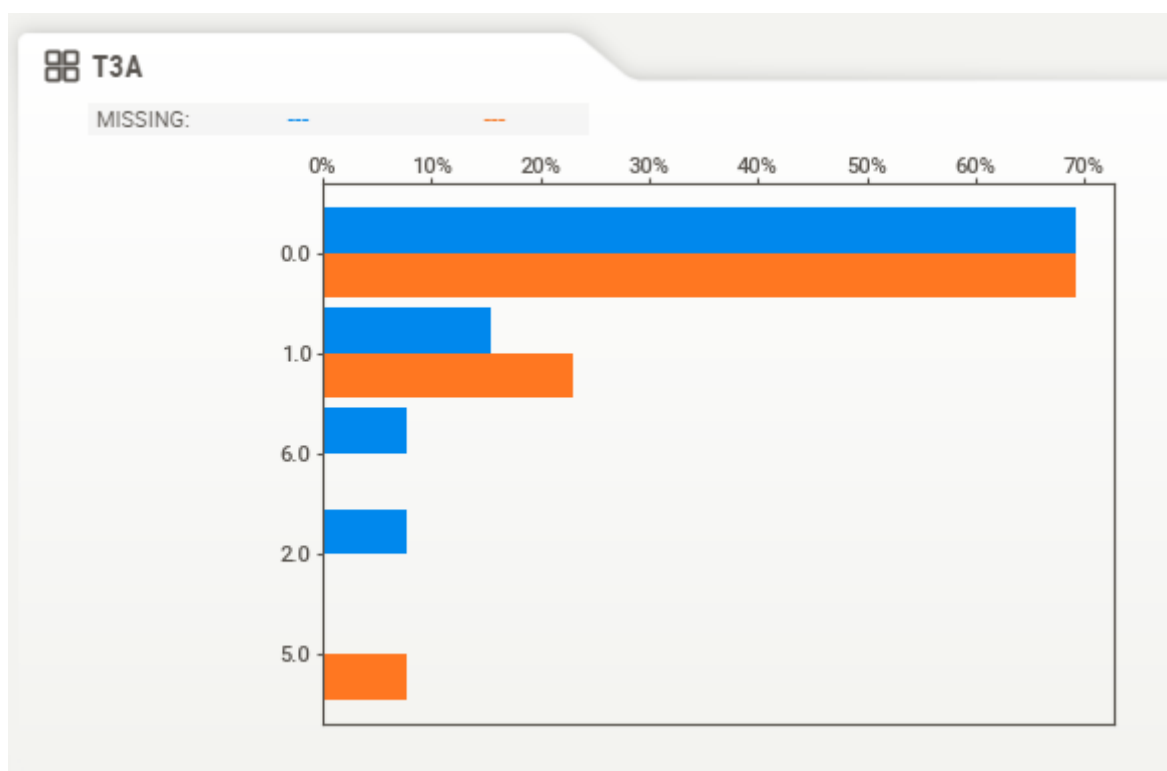


Figura 11. Análisis de T3A entre partidos del Grupo Constructalia Tacoronte y RCNT

## 5. Evaluación de los resultados por parte de los expertos

Se puede observar como un mayor número de jugadores no anotaron en el primer partido,

mientras que en el segundo partido la anotación fue más repartida.

Respecto a los T3I y T3A no hay gran variación entre un partido y otro lo que hace indicar que están en la media de su temporada.

#### 6. Aplicación individual/grupal

Realizar entrenamientos en los que se practiquen tanto las situaciones individuales como colectivas, buscando repartir la anotación del equipo en distintos jugadores.

Efectuar ejercicios de lanzamiento de 3 puntos buscando mejorar los porcentajes del equipo.

#### 7. Comparación con los datos de origen

Para finalizar se comparan los datos originales con los datos obtenidos después de realizar las sesiones de entrenamientos permitiendo ver si se ha realizado alguna mejora en su rendimiento.

### ***Caso 4: Estadísticas locales, relación de T1% y Pts individuales***

#### 1. Recopilación de datos estadísticos

Para este caso, los datos han sido proporcionados por un entrenador del Cajasiete Canarias, contamos con las estadísticas de todos los partidos de la temporada.

#### 2. Preprocesamiento y limpieza del conjunto de datos

Los datos se han transformado desde .xlsx a .csv. Se han eliminado filas y columnas innecesarias y se han realizado modificaciones en su conjunto añadiendo valores que estaban vacíos y cambiando el tipo de algunos datos.

#### 3. Análisis de los datos para evaluar el rendimiento

Se va a analizar la relación entre el porcentaje de tiros libres (T1%) y los puntos anotados (PTS) de forma individual.

#### 4. Visualización de los resultados

La visualización se ha realizado con la librería matplotlib.pyplot.

Los datos representados corresponden a la relación entre el porcentaje de tiros libres anotados (eje x) y los puntos anotados (eje y) de manera individual durante toda la temporada.

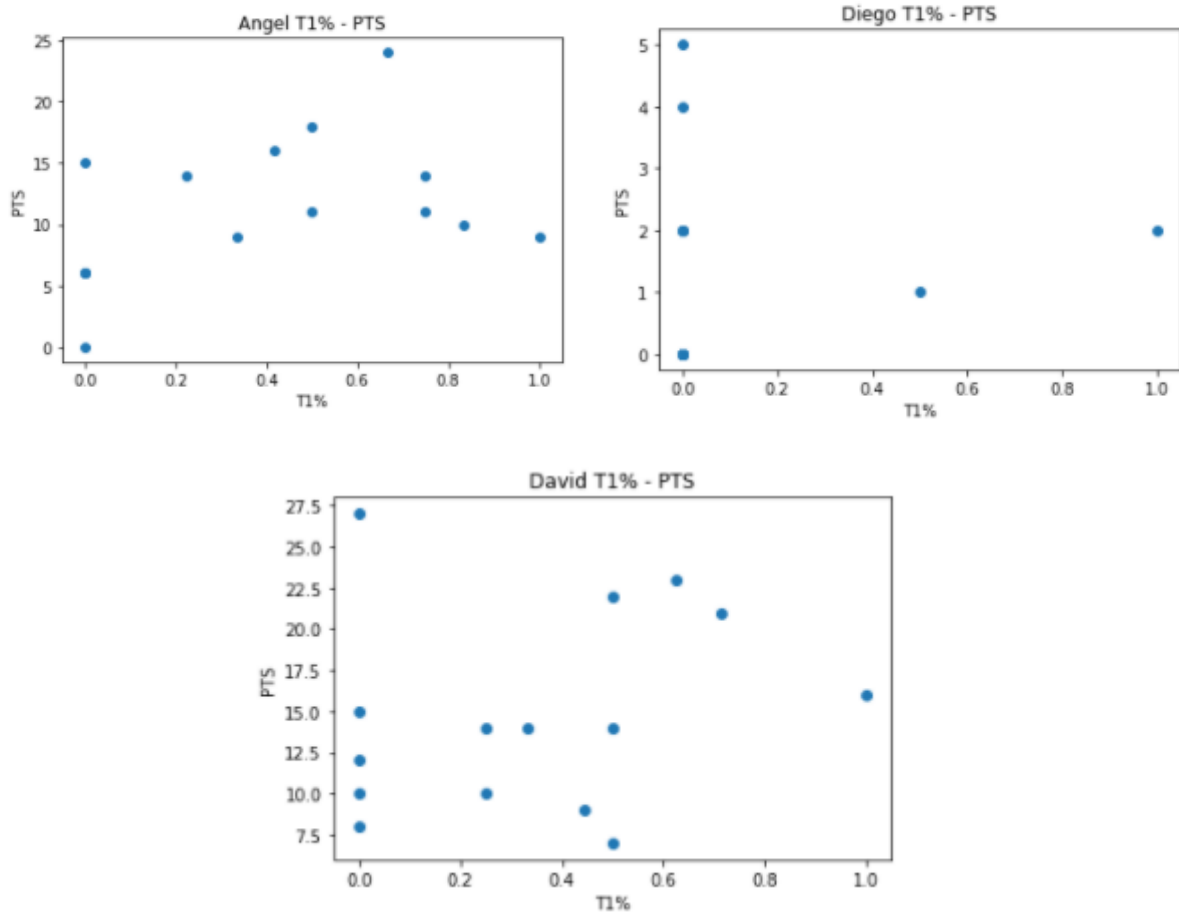


Figura 12. Relación entre T1% y PTS para los jugadores Diego, Angel y David del Cajasiete Canarias

## 5. Evaluación de los datos por parte de los expertos

Ángel: necesita mejorar los porcentajes para incrementar su nivel de anotación.

Diego: necesita mejorar los porcentajes, se observa un mayor número de partidos con un 0%.

David: sus porcentajes son bajos a pesar de tener una gran capacidad anotadora, necesita mejorar los tiros libres para sacar mayor rendimiento a su juego.

## 6. Aplicación individual/grupal.

Realizar ejercicios de tiros libres, en diferentes situaciones, perfeccionando los detalles y corrigiendo errores.

## 7. Comparación con los datos de origen

Para finalizar se comparan los datos originales con los datos obtenidos después de realizar las sesiones de entrenamientos permitiendo ver si se ha realizado alguna mejora en su rendimiento.

## **Capítulo 5 CONCLUSIONES Y LÍNEAS FUTURAS**

### **5.1 CONCLUSIONES**

Con el cierre de este proyecto hemos llegado a las siguientes conclusiones:

Se obtuvo una metodología de trabajo con datos estadísticos para facilitar el estudio del rendimiento de los jugadores / equipos de baloncesto, se desarrollaron scripts para facilitar su aplicación y se probó con datos reales de nivel profesional, amateur y de formación.

En un primer momento se han analizado datos de anotación de jugadores en determinadas posiciones de la cancha, facilitando la visualización de las zonas con mayor o menor anotación y permitiendo tomar la decisión de las zonas en las que hacer mayor trabajo técnico durante los entrenamientos.

La utilización de elementos gráficos es imprescindible para la interpretación del análisis. Como se ha observado en los gráficos generados desde las estadísticas del equipo Grupo Constructalia Tacoronte, se puede realizar una comparación entre distintas jornadas de una forma sencilla y rápida, obteniendo resultados grupales e individuales con los que trabajar con vistas al futuro desarrollo de las aptitudes

En equipos de formación debería ser una herramienta esencial para el desarrollo de los jugadores.

Todo equipo, independientemente de la categoría, debería tener acceso a herramientas para el análisis del rendimiento porque no es necesario disponer de un presupuesto elevado.

Por último, es indispensable realizar un trabajo externo en entrenamientos y partidos para poder aplicar los resultados de la metodología, sin el trabajo individual/grupal sobre el análisis naturalmente los resultados serán inexistentes.

### **5.2 LÍNEAS FUTURAS**

Se han establecido las bases para la continuación del desarrollo de la metodología de análisis del rendimiento dentro del baloncesto.

El trabajo a continuar se centra en la incorporación de distintas métricas que proporcionan valor a su análisis y ayude a estudiar el rendimiento de los jugadores.

De igual forma, se puede realizar el desarrollo de un servicio que facilite la recolección de datos, su análisis y posterior visualización de los resultados.

## **Capítulo 6 SUMMARY AND CONCLUSIONS**

### **6.1 CONCLUSIONS**

With the closure of this project we have reached the following conclusions:

A working methodology with statistical data was obtained to facilitate the study of the performance of basketball players / teams, scripts were developed to facilitate its application and it was tested with real data at the professional, amateur and training levels.

At first, annotation data of players in certain positions on the court has been analyzed, facilitating the visualization of the areas with more or less annotations and allowing the decision of the areas in which to do more technical work during training.

The use of graphic elements is essential for the interpretation of the analysis. As has been observed in the graphs generated from the statistics of the Grupo Constructalia Tacoronte team, a comparison can be made between different days in a simple and fast way, obtaining group and individual results with which to work with a view to the future development of skills.

In training teams it should be an essential tool for the development of the players.

Every team, regardless of category, should have access to tools for performance analysis because you don't need a large budget.

Finally, it is essential to carry out external work in training sessions and games in order to apply the results of the methodology, without individual / group work on the analysis, the results will naturally be non-existent.

### **6.2 FUTURE LINES**

The foundations have been laid for the further development of the performance analysis methodology within basketball.

The work to be continued focuses on incorporating different metrics that provide value to your analysis and help you study player performance.

Similarly, the development of a service that facilitates the collection of data, its analysis and subsequent visualization of the results can be carried out.

## Capítulo 7 PRESUPUESTO

En este capítulo se presenta una estimación sobre el coste de la realización de este proyecto. También se presenta un coste hipotético de la implementación de este modelo en un entorno real.

### 7.1 PRESUPUESTO DE RECURSOS HUMANOS

En este apartado se muestra de forma desglosada el coste de los recursos humanos que ha sido invertido en el desarrollo del proyecto.

Tarea	Cantidad de horas empleadas	Coste bruto por hora	Coste total
Estudios relacionados al Ciencia de datos / Inteligencia Artificial	20	14€	480€
Estudio de las diferentes arquitecturas	15	14€	210€
Búsqueda y análisis de los conjuntos de datos	18	14€	252€
Procesamiento y transformación de los datos	10	14€	140€
Visualización de los datos	8	14€	112€
Aprendizaje de las librerías y módulos	25	14€	350€
Elección y creación del modelo	20	14€	280€
Documentación de la aplicación	10	14€	140€
Total:	126 horas		1973€

Tabla 1. Presupuesto de Recursos Humanos

### 7.2 PRESUPUESTO MATERIAL

En este apartado se calculan los costes de los distintos componentes necesarios para el desarrollo de la aplicación.

Elemento	Coste total
Portátil	950€
Librerías / Módulos externos	0€
IDE / Editor	0€
Total:	950€

Tabla 2. Presupuesto de los materiales necesarios para el supuesto caso

### 7.3 PRESUPUESTO FINAL

El presupuesto final teniendo en cuenta el presupuesto del personal humano y utilizando el presupuesto de los materiales necesarios para la realización de proyecto, sería el siguiente:

<b>Elemento</b>	<b>Coste total</b>
Coste humano	1973€
Coste de los materiales	950€
Total:	2923€

*Tabla 3. Presupuesto final.*



## Capítulo 8 APÉNDICES

El código desarrollado está disponible en el siguiente repositorio:

<https://github.com/seerm22/TFG>

### 8.1 SCRIPT PARA EXTRAER LOS PTS ANOTADOS EN CADA ZONA DE LA CANCHA Y REPRESENTARLOS GRÁFICAMENTE

```
# NBA_API
from nba_api.stats.endpoints import shotchartdetail
from nba_api.stats.static.players import get_active_players
from nba_api.stats.static.teams import get_teams
from nba_api.stats.static import players
#JSON
import json
#PANDAS
import pandas as pd
#MATPLOTLIB
import matplotlib as mpl
import matplotlib.pyplot as plt
# Obtener el listado de jugadores en activo de la NBA
players = get_active_players()
# Obtener el listado de equipos de la NBA
teams = get_teams()
# Funciones para extraer los id según nombre de jugador y franquicia
def get_player_id(first, last):
    for player in players:
        if player['first_name'] == first and player['last_name'] == last:
            return player['id']
    return -1
def get_team_id(team_name):
    for team in teams:
        if team['full_name'] == team_name:
            return team['id']
    return -1
luka_doncic_pts_rs = shotchartdetail.ShotChartDetail(team_id=get_team_id('Dallas Mavericks'),
player_id=get_player_id('Luka', 'Doncic'),
context_measure_simple=('PTS'),
season_type_all_star='Regular Season',
season_nullable = '2020-21'
)
doncic_shot_data = json.loads(luka_doncic_pts_rs.get_json())
relevant_data = doncic_shot_data['resultSets'][0]
headers = relevant_data['headers']
rows = relevant_data['rowSet']
# Create pandas DataFrame
doncic_data = pd.DataFrame(rows)
doncic_data.columns = headers
# LUKA DONCIC CLUTCH TIME (30 S) PLAYOFFS
```

```

luka_doncic_pts_clutch = shotchartdetail.ShotChartDetail(team_id=get_team_id('Dallas
Mavericks'),
player_id=get_player_id('Luka', 'Doncic'),
context_measure_simple=('PTS'),
season_type_all_star='Playoffs',
clutch_time_nullable='Last 30 Seconds',
season_nullable = '2020-21'
)
doncic_clutch_data = json.loads(luka_doncic_pts_clutch.get_json())
#Clutch
relevant_data = doncic_clutch_data['resultSets'][0]
headers = relevant_data['headers']
rows = relevant_data['rowSet']
doncic_clutch = pd.DataFrame(rows)
doncic_clutch.columns = headers

```

## 8.2 ANÁLISIS CON SWEETVIZ

```

import pandas as pd
import sweetviz as sv
df = pd.read_csv('../static/canarias_infantil_masculino/csv/CE-canarias-valencia.csv')
canarias = sv.analyze(df)
canarias.show_html('../output/canarias-valencia.html')
rcnt_tacoronte = pd.read_csv('../static/tacoronte_primera_autonomica/csv/rcnt-tacoronte.csv')
rcnt_tacoronte['T1%'] = pd.to_numeric(df['T2%'], errors='coerce')
rcnt_tacoronte['T2%'] = pd.to_numeric(df['T2%'], errors='coerce')
rcnt_tacoronte['T3%'] = pd.to_numeric(df['T2%'], errors='coerce')
rcnt_tacoronte.fillna(0, inplace= True)
tacoronte_rcnt = pd.read_csv('../static/tacoronte_primera_autonomica/csv/tacoronte-rcnt.csv')
tacoronte_rcnt['T1%'] = pd.to_numeric(df['T2%'], errors='coerce')
tacoronte_rcnt['T2%'] = pd.to_numeric(df['T2%'], errors='coerce')
tacoronte_rcnt['T3%'] = pd.to_numeric(df['T2%'], errors='coerce')
tacoronte_rcnt.fillna(0, inplace= True)
report = sv.compare(rcnt_tacoronte,tacoronte_rcnt)
report.show_html('../output/report.html')

```

## 8.3 ANÁLISIS INDIVIDUAL T1% - PTS

```

import pandas as pd
from matplotlib import pyplot as plt
import glob
l = [pd.read_csv(filename) for filename in
glob.glob("../static/canarias_infantil_masculino/csv/*.csv")]
df =pd.concat(l, axis=0)
df['T1%'] = df['T1%'].str.replace(',', '.').astype(float)
df['T2%'] = df['T2%'].str.replace(',', '.').astype(float)
df['T3%'] = df['T3%'].str.replace(',', '.').astype(float)

```

```
df.fillna(0, inplace= True)
angel = df[df.Nombre == 'Angel']
diego = df[df.Nombre == 'Diego']
david = df[df.Nombre == 'David']
plt.plot(angel['T1%'],angel.Pts,'o')
plt.title('Angel T1% - PTS')
plt.xlabel('T1%')
plt.ylabel('PTS')
plt.show()
plt.plot(diego['T1%'],diego.Pts,'o')
plt.title('Diego T1% - PTS')
plt.xlabel('T1%')
plt.ylabel('PTS')
plt.show()
plt.plot(david['T1%'],david.Pts,'o')
plt.title('David T1% - PTS')
plt.xlabel('T1%')
plt.ylabel('PTS')
plt.show()
```

## REFERENCIAS

[1] LI, Bin; XU, Xinyang. Application of Artificial Intelligence in Basketball Sport. Journal of Education, Health and Sport, [S.l.], v. 11, n. 7, p. 54-67, jul. 2021. ISSN 2391-8306. Available at: <https://apcz.umk.pl/czasopisma/index.php/JEHS/article/view/JEHS.2021.11.07.005>.

[2] Chen, W.-J.; Jhou, M.-J.; Lee, T.-S.; Lu, C.-J. Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association. Entropy 2021, 23, 477. <https://doi.org/10.3390/e23040477>

[3] Dembe Koi Stephanos, Ghaith Husari, Brian T. Bennett, and Emma Stephanos. 2021. Machine learning predictive analytics for player movement prediction in NBA: applications, opportunities, and challenges. Proceedings of the 2021 ACM Southeast Conference. Association for Computing Machinery, New York, NY, USA, 2–8. DOI:<https://doi.org/10.1145/3409334.3452064>

[4] Zafar Mahmood, Ali Daud, Rabeeh Ayaz Abbasi, Using machine learning techniques for rising star prediction in basketball, Knowledge-Based Systems, Volume 211, 2021, 106506, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2020.106506>. <https://www.sciencedirect.com/science/article/pii/S0950705120306353>

[5] Emad Abdelrasoul, Islam Mahmoud, Pro Stergiou, Larry Katz, The Accuracy of a Real Time Sensor in an Instrumented Basketball, Procedia Engineering, Volume 112, 2015, Pages 202-206, ISSN 1877-7058, <https://doi.org/10.1016/j.proeng.2015.07.200>. (<https://www.sciencedirect.com/science/article/pii/S1877705815014496>

[6] Vangelis Sarlis, Vasilis Chatziilias, Christos Tjortjis, Dimitris Mandalidis, A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance, Information Systems, Volume 99, 2021, 101750, ISSN 0306-4379, <https://doi.org/10.1016/j.is.2021.101750>. (<https://www.sciencedirect.com/science/article/pii/S030643792100020X>

[7] Van Rossum, G. (2007, June). Python Programming Language. In USENIX annual technical conference (Vol. 41, p. 36). [https://thereaderwiki.com/en/Python\\_\(programming\\_language\)](https://thereaderwiki.com/en/Python_(programming_language))

[8] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... &

Willing, C. (2016). Jupyter Notebooks-a publishing format for reproducible computational workflows (Vol. 2016, pp. 87-90).<https://books.google.es/books?id=Lgy3DAAAQBAJ>

[9] Swar. Nba\_api Repository. [https://github.com/swar/nba\\_api](https://github.com/swar/nba_api)

[10] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03),90-95.[Matplotlib: A 2D Graphics Environment](#)

[11] McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, 14(9),1-9  
[https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011\\_submission\\_9.pdf](https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf)

[12] Elansary, M. (2021). Data wrangling & preparation automation.[shorturl.at/fpOTU](https://shorturl.at/fpOTU)