



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

## Trabajo de Fin de Grado

---

**Ciencia de datos con Python en su  
aplicación a la práctica del fútbol.**

*Data science with Python in its application to soccer.*

Noah Rafael Sanchez Geurts

---

La Laguna, 8 de julio de 2022

D. **José Andrés Moreno Pérez**, con N.I.F. 42.935.437-A, Catedrático 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**

Que la presente memoria titulada:

*"Ciencia de datos con Python en su aplicación a la práctica del fútbol"*

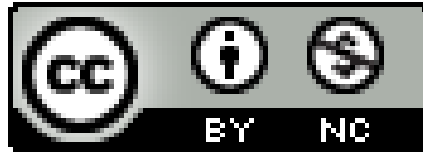
ha sido realizada bajo su dirección por D. **Noah Rafael Sánchez Geurts**, con N.I.F. 45.939.668-C.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 8 de julio de 2022

# Agradecimientos

A mi familia y compañeros por la motivación y los consejos que me han dado durante la carrera. A mi tutor de TFG por todos los consejos y apoyo durante la elaboración de este TFG.  
Y en especial a mi pareja por ser mi motivación cada día y apoyarme en este trabajo.

# Licencia



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

## **Resumen**

*El objetivo de este trabajo es la aplicación de ciencia de datos con Python en el sector deportivo, concretamente en la práctica del fútbol.*

*A través del uso de distintas librerías de Python como Pandas, Matplotlib y Sweetviz, junto con diferentes técnicas estadísticas básicas y avanzadas, así como aplicación de modelos matemáticos, se propone definir una metodología para realizar una serie de estudios sobre ciencia de datos aplicada a la práctica del fútbol, con la finalidad de indagar con detalle sobre las lesiones, factores de riesgo y anatomía de los jugadores de un equipo de fútbol, así como mediante diversas técnicas de análisis de datos como data-tracking se pretende profundizar en los diversos planteamientos estratégicos o disposiciones tácticas de los equipos durante un partido de fútbol, además de obtener información del rendimiento de los jugadores individualmente, o en su conjunto, obtener diferentes estadísticas que puedan afectar al rendimiento de los jugadores y del equipo. Todo esto con el propósito de transformar esta información a conocimiento para que el equipo técnico pueda aprovecharlo y mejorar el rendimiento del equipo, así como agilizar la toma de decisiones.*

**Palabras clave:** Datos, Python, Análisis, Estadística, Algoritmo, Ciencia de Datos, Fútbol

## **Abstract**

*The objective of this work is the application of data science with Python in sports, particularly in soccer.*

*Through the use of different Python libraries such as Pandas, Matplotlib and Sweetviz, together with different basic and advanced statistical techniques, as well as the application of mathematical models, it is proposed to define a methodology to carry out a series of studies on data science applied to the practice of soccer, in order to investigate in detail the injuries, risk factors and anatomy of the players of a soccer team, as well as through various data analysis techniques such as data-tracking, it is intended to delve into the various strategic approaches or team tactics during a football match, in addition to obtaining information on the performance of the players individually, or as a whole, obtaining different statistics that may affect the performance of the players and the team. All this with the purpose of transforming this information into knowledge so that the technical team can take advantage of it and improve the team's performance, as well as speed up decision-making.*

**Keywords:** Data, Python, Analysis, Statistics, Algorithm, Data Science, Soccer

# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes</b>	<b>3</b>
2.1. Estado del Arte . . . . .	3
2.2. Tecnologías, Herramientas, Librerías . . . . .	4
2.3. Bases de Datos . . . . .	5
2.3.1. Lesiones . . . . .	5
2.3.2. Data-tracking . . . . .	6
2.4. Metodología . . . . .	6
<b>3. Objetivos</b>	<b>7</b>
3.1. Objetivo general . . . . .	7
3.2. Objetivos específicos . . . . .	7
3.2.1. Estudio de lesiones . . . . .	7
3.2.2. Estudio de data-tracking . . . . .	7
<b>4. Desarrollo</b>	<b>8</b>
4.1. Entorno de desarrollo . . . . .	8
4.2. Caso 1: Lesiones . . . . .	8
4.2.1. Visualizaciones Básicas . . . . .	9
4.2.2. Modificación de los datos . . . . .	9
4.2.3. Factores de Riesgo . . . . .	12
4.3. Caso 2: Data-tracking . . . . .	15
4.3.1. Modificación inicial de los datos . . . . .	15
4.3.2. Visualizaciones básicas . . . . .	17
4.3.3. Calculando velocidad . . . . .	18
4.3.4. Estadísticas de equipo . . . . .	20
4.3.5. Visualizaciones avanzadas . . . . .	22
<b>5. Conclusiones y líneas futuras</b>	<b>26</b>
5.1. Conclusiones . . . . .	26
5.2. Líneas futuras . . . . .	26
<b>6. Summary and Conclusions</b>	<b>27</b>
<b>7. Presupuesto</b>	<b>28</b>
7.1. Presupuesto para el personal . . . . .	28
7.2. Presupuesto material . . . . .	28
7.3. Presupuesto Total . . . . .	29

<b>A. Algoritmos del caso de lesiones</b>	<b>30</b>
A.1. Algoritmo para separar el vector de lesiones en diferentes celdas . . . . .	30
A.2. Algoritmo para visualizar las lesiones de cada equipo . . . . .	30
A.3. Algoritmo para mostrar las lesiones combinando factores de riesgo . . . . .	31
<b>B. Algoritmos del caso de data-tracking</b>	<b>32</b>
B.1. Algoritmo para modificar los datos de las posiciones . . . . .	32
B.2. Algoritmo para calcular la velocidad de los jugadores . . . . .	33
B.3. Algoritmo para mostrar los jugadores en el campo . . . . .	33
<b>Siglas</b>	<b>35</b>



# Índice de Figuras

2.1. Datos básicos de los jugadores . . . . .	5
2.2. Factores de riesgo de los jugadores . . . . .	5
2.3. Tabla de los eventos del partido . . . . .	6
4.1. Número de lesiones por equipo (Previas) . . . . .	9
4.2. Número de lesiones por equipo (Actuales) . . . . .	9
4.3. Número de lesiones por equipo agrupadas por posición de los jugadores (Previas) . . . . .	10
4.4. Número de lesiones por equipo agrupadas por posición de los jugadores (Actuales) . . . . .	10
4.5. Total lesiones por zona (Previas) . . . . .	11
4.6. Total lesiones por zona (Actuales) . . . . .	11
4.7. Total lesiones por parte (Previas) . . . . .	11
4.8. Total lesiones por parte (Actuales) . . . . .	11
4.9. Modificación de la tabla a partir del método explode . . . . .	12
4.10 Jugadores con V/V Calcáneo . . . . .	13
4.11 Jugadores con 1º dedo rígido . . . . .	13
4.12 Jugadores con V/V Rodilla . . . . .	13
4.13 Jugadores con Glúteo Medio . . . . .	13
4.14 Lesiones con V/V Calcáneo . . . . .	15
4.15 Lesiones con V/V Rodilla . . . . .	15
4.16 Lesiones con Gl. Medio . . . . .	16
4.17 Lesiones con 1º dedo rígido . . . . .	16
4.18 Formato inicial de los datos de posiciones de los jugadores . . . . .	16
4.19 Formato de los datos de las posiciones modificados . . . . .	16
4.20 Gráfico del terreno de juego . . . . .	17
4.21 Primer evento del partido . . . . .	17
4.22 Posición inicial de los jugadores y el balón . . . . .	18
4.23 Posición inicial con nombres . . . . .	18
4.24 Velocidad de los jugadores . . . . .	19
4.25 Distancia recorrida por cada jugador del equipo local (Km) . . . . .	19
4.26 Distancia recorrida por jugador clasificada por velocidad . . . . .	20
4.27 Posesión de balón . . . . .	21
4.28 Pases acertados . . . . .	21
4.29 Pases completados del equipo visitante . . . . .	22
4.30 Pases completados del equipo local . . . . .	22
4.31 Tiros totales . . . . .	22
4.32 Tiros a puerta . . . . .	22
4.33 Tiros del equipo visitante . . . . .	22

4.34	Tiros del equipo local . . . . .	22
4.35	Ejemplo de Pitch Control . . . . .	23
4.36	Asistencia del primer gol . . . . .	24
4.37	Asistencia del segundo gol . . . . .	24
4.38	Asistencia del tercer gol . . . . .	24
4.39	Asistencia del cuarto gol . . . . .	24
4.40	Clasificación de los pases del equipo local . . . . .	25
4.41	Clasificación de los pases del equipo visitante . . . . .	25

# Índice de Tablas

4.1. Tabla de la información básica del jugador . . . . .	8
4.2. Tabla de los distintos factores de riesgo . . . . .	9
7.1. Presupuesto de las tareas . . . . .	28
7.2. Presupuesto material . . . . .	29
7.3. Presupuesto material . . . . .	29

# Capítulo 1

## Introducción

En este capítulo aportaremos una breve introducción del tema a tratar, la ciencia de datos y sus aplicaciones en la práctica del fútbol.

La Ciencia de Datos es un subconjunto de la Inteligencia Artificial (IA) que se refiere más a las áreas superpuestas de las estadísticas, los métodos científicos y el análisis de datos.

Tiene como objetivo extraer conocimiento de grandes bases de datos. Incluye varios componentes como preparación de datos, estadística, aprendizaje automático y visualización de datos. La preparación de datos incluye el proceso de inspeccionar, limpiar y transformar datos con el objetivo de seleccionar información útil.

La ciencia de datos o el análisis de datos se ha convertido en un componente esencial de las operaciones organizacionales. La capacidad de recopilar e interpretar conjuntos de datos significativamente grandes han proporcionado una gran cantidad de conocimientos para guiar a los tomadores de decisiones en todas las facetas de la sociedad, desde la optimización de cadenas de abastecimiento, automatización de procesos, monitoreo y control de calidad, sistemas de mantenimiento predictivo, etc.

Gracias al paso del tiempo y al avance de las tecnologías estas técnicas se han conseguido aplicar en el ámbito deportivo [10] utilizando grandes datos dentro y fuera del campo para agilizar la toma de decisiones.

Actualmente se aplican técnicas de seguimiento de datos, o data tracking, con la ayuda de diferentes cámaras localizadas en todo el campo, además con la implementación de la IA, para poder conocer el posicionamiento de cada jugador, así como del balón en todo momento. Con esta técnica tan compleja podemos conseguir las estadísticas (pases, tiros, paradas, saques de banda, etc.) de cada partido, o de cada sesión de entrenamiento en tiempo real, las cuales utilizan los cuerpos técnicos de cada equipo para poder mejorar el rendimiento de sus jugadores.

En el fútbol no solo se aplican estrategias analíticas dentro del campo, también son utilizadas en los aspectos financieros, como fichajes, salarios, marketing, etc. [2]. Por otro lado también se utilizan estas técnicas en el ámbito social, mas enfocado a las redes sociales, espectadores, audiencia televisiva, *merchandising*, etc.

En este trabajo, junto a herramientas de Python para el desarrollo de ciencia de datos, realizaremos una serie de estudios de ciencia de datos sobre riesgo y prevención de lesiones, análisis estadísticos de equipos y análisis de rendimiento de jugadores mediante técnicas de data tracking.

El trabajo se divide en seis capítulos, introducción breve del tema a tratar, antecedentes, donde hablamos de la situación actual de la ciencia de datos aplicada al fútbol, tecnologías empleadas, bases de datos, etc. Luego entraremos en el detalle del desarrollo realizado,

visualizaciones y resultados. Y por último aportaremos las conclusiones obtenidas del estudio.

# Capítulo 2

## Antecedentes

En este capítulo indagaremos sobre la actualidad de las aplicaciones de la ciencia de datos en los distintos deportes, y en el fútbol. Además, introduciremos las diferentes tecnologías empleadas, las diferentes bases de datos que hemos utilizado, así como una breve explicación de la metodología de trabajo.

### 2.1. Estado del Arte

En la actualidad, una gran parte de las grandes empresas tienen a su servicio equipos expertos en ciencia de datos, los cuales, con la ayuda de técnicas analíticas como el análisis predictivo, visualización de datos, Big Data (BD) e IA consiguen mejorar el rendimiento de sus procesos o ganancias.

Lo mismo sucede con las grandes instituciones deportivas, las cuales utilizan estas herramientas mencionadas para obtener el máximo rendimiento, en este caso, de sus jugadores, ventas, contrataciones, etc. La NBA ha estado usando estadísticas con tanta frecuencia [1] que podría superar a Major League Baseball (MLB) en el uso de datos, que fue una de las primeras ligas deportivas de Estados Unidos en usar datos para encontrar información y patrones ocultos para beneficiar a su equipo. "Money Ball", tanto la película como el libro ofrecen una buena descripción de los eventos y el uso de datos que llevaron a un equipo de bajos recursos (Oakland Athletics) a ganar muchos partidos con jugadores que estaban infravalorados. El fútbol tampoco se queda atrás, actualmente es uno de los deportes que más aprovecha estas herramientas [8].

La Premier League (máxima categoría del fútbol inglés) como La Liga Santander, (máxima categoría del fútbol español) son las ligas que más aprovechan estas herramientas tan avanzadas, por ejemplo La Liga Santander, junto con la empresa tecnológica Microsoft han desarrollado Beyond Stats (<https://www.laliga.com/beyondstats>), un portal web abierto al público. Este ofrece un gran conjunto de datos dinámicos a través de Machine Learning (ML), BD e IA para todo el mundo que quiera consultarlo, esto incluye también a los mismos clubes que quieran sacarle provecho para analizar sus partidos, jugadas, estadísticas, etc.

Además, gracias a la IA contamos con herramientas que nos ayudan a hacer el juego más justo, como son el V.A.R o videoarbitraje que facilita al árbitro la toma de decisiones, o el Ojo de Halcón, que mediante reconocimiento de imágenes se le comunica al árbitro si el balón ha entrado por completo en la portería. Este mismo 2022 se implantará una nueva herramienta llamada Fuera de juego semiautomático, con el que se podrá detectar si un atacante está en posición de fuera de juego de forma precisa en segundos. Esto es

gracias a 12 cámaras situadas en el estadio que mediante una IA [5] y el reconocimiento de imágenes reconocerá la posición de todos los futbolistas en el terreno de juego para determinar si es fuera de juego o no.

Otro sector que utiliza en gran medida estas estadísticas es el de las apuestas deportivas, de hecho, son su núcleo de funcionamiento. Cuando muestran cuánto dinero se puede ganar apostando por un resultado, no se hace de forma aleatoria. Estas cifras tienen como base la información obtenida con las estadísticas, BD y análisis de partidos anteriores, así como de los encuentros que se han disputado entre ambos equipos a través de la historia. Con estos números se calculan las probabilidades de que se dé un resultado concreto, y cuanto menores sean estas, mayor es la cantidad de dinero que se puede conseguir apostando.

En realidad, las estadísticas son matemáticas, y como tales se pueden emplear para analizar, mejorar o incluso predecir lo que suceda en el futuro para ser cada vez más expertos en el deporte rey.

## 2.2. Tecnologías, Herramientas, Librerías

La ciencia de datos es uno de los puntos más fuertes dentro del campo de la innovación deportiva, nos aporta una gran cantidad de información útil con la que poder mejorar nuestro rendimiento. En esta se incluye una serie de campos como la IA, BD, ML, la estadística, análisis de datos, visualización de datos, etc. Además, tiene que pasar por una serie de procesos de manipulación de datos, donde el usuario limpia o modifica los datos que tiene para transformarlos en conocimiento.

Para poder realizar todos estos procesos se utilizó Python, uno de los lenguajes de programación más solicitados actualmente, además, es junto con R, el lenguaje más recomendado para hacer este tipo de análisis estadísticos y visualizaciones de datos.

Las herramientas que utilizamos para crear un entorno de desarrollo fueron Visual Studio Code, a través de una máquina virtual de Linux, y también a través de Google Colab [9], una herramienta que permite a cualquier usuario escribir y ejecutar código de Python en el navegador. Diseñado específicamente para realizar tareas de aprendizaje automático, análisis de datos y educación. Gracias a esta herramienta podemos almacenar todo el código en la nube sin perderlo y pudiendo acceder a él desde cualquier dispositivo.

Para realizar este estudio se utilizaron una serie de librerías o paquetes de Python con las que poder efectuar los análisis de los datos así como las visualizaciones:

- **NumPy**. Paquete que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas de alto nivel.
- **Pandas** [4]. Biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos en Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales.
- **Matplotlib** [3]. Biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación Python
- **csv**. Biblioteca que implementa clases y funciones para leer y escribir datos en formato CSV.

- **sweetvix**. Biblioteca de análisis de datos exploratorios con el objetivo de descubrir tipos de datos, estadísticas, distribución de valores, correlaciones, etc.

## 2.3. Bases de Datos

### 2.3.1. Lesiones

Como mencionamos anteriormente, en este trabajo hemos realizado una serie de estudios, uno de análisis de lesiones, en este estudio contamos con la colaboración de dos fisioterapeutas de tres clubes de las categoría amateur del fútbol español, todo esto con previo consentimiento de los jugadores y respetando la ley de protección de datos. Los datos que nos proporcionan son sobre la información básica del jugador (Figura 2.1), lesiones de las temporadas pasada y actual, así como de los factores de riesgo de cada uno de ellos (Figura 2.2). Los datos fueron proporcionados en formato Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
	Equipo	Jugador	Edad	Altura	Lesiones Previas	Total Lesiones Previas	Lesiones Actuales	Total Lesiones Actuales	Operaciones	Diabetes	Corrección ort.	Escoliosis
2	Padre Anchieta	Jugador1	22	181	obillo-Tobillo-PIE ; Esguince de tobillo-	4	o tobillo-Tobillo-PIE ; Tenosinovitis-Rodi	2	Hernia inguinal	0	0	1
3	Padre Anchieta	Jugador2	22	185	Molestia de psoas-Psoas-CADERA	1	ura de isquiotibiales-Isquiosurales-MUS	1	No	0	0	0
4	Padre Anchieta	Jugador3	20	180	uxación de hombro-Hombro-HOMBR	1	ractura de aductores-Aductores-MU	1	No	0	0	0
5	Padre Anchieta	Jugador4	26	175	No-Sin lesión	0	No-Sin lesión	0	No	0	0	1
6	Padre Anchieta	Jugador5	22	170	Esguince de tobillo-Tobillo-PIE	1	billo-PIE ; Contratura de isquiotibiales-	2	No	0	0	0
7	Padre Anchieta	Jugador6	22	182	utura de cuádriceps-Cuádriceps-MUSL	1	ura de isquiotibiales-Isquiosurales-MUS	1	No	0	0	0
8	Padre Anchieta	Jugador7	23	180	lo-Tobillo-PIE ; Esguince de tobillo-Tob	7	utura de aductores-Aductores-MUSL	1	No	0	0	1
9	Padre Anchieta	Jugador8	25	160	No-Sin lesión	0	pbrecarga de gemelos-Gemelos-PIERN	1	No	0	0	0
10	Padre Anchieta	Jugador9	22	192	No-Sin lesión	0	iquiosurales-MUSLO ; Rotura de cuádr	2	No	0	0	0
11	Padre Anchieta	Jugador10	21	176	IA-Rodilla-PIERNA ; Esguince de tobill	2	No-Sin lesión	0	Rec Lig. Cruz.	0	1	0
12	Padre Anchieta	Jugador11	18	174	No-Sin lesión	0	No-Sin lesión	0	No	0	0	0
13	Padre Anchieta	Jugador12	20	186	No-Sin lesión	0	ica-Muñeca-MANO ; Fractura de meñiq	2	No	0	0	0
14	Padre Anchieta	Jugador13	25	180	Lumbalgia-Lumbar-TRONCO	1	No-Sin lesión	0	No	0	0	0
15	Padre Anchieta	Jugador14	18	174	No-Sin lesión	0	Esguince de tobillo-Tobillo-PIE	1	No	0	0	0
16	Padre Anchieta	Jugador15	21	177	No-Sin lesión	0	No-Sin lesión	0	No	0	0	0
17	Padre Anchieta	Jugador16	20	183	No-Sin lesión	0	No-Sin lesión	0	No	0	0	0
18	Marino	Jugador17	21	179	infractura de gemelos-Gemelos-PIERN	1	nusión de cuádriceps-Cuádriceps-MUS	1	No	0	0	0
19	Marino	Jugador18	25	185	ibiema de pata de ganso-Rodilla-PIER	1	ganso-Rodilla-PIERNA ; Rotura de BF-I	2	Labrum	0	0	1
20	Marino	Jugador19	23	174	No-Sin lesión	0	No-Sin lesión	0	No	0	0	0
21	Marino	Jugador20	21	183	leolo y rotula-Rodilla-PIERNA ; Esguin	2	IA ; Rotura de gemelos-Gemelos-PIERN	3	No	0	0	0
22	Marino	Jugador21	18	175	No-Sin lesión	0	No-Sin lesión	0	No	0	0	1
23	Marino	Jugador22	25	180	ón de musculos peroneos-Peroneos-P	1	No-Sin lesión	0	and. peroneos dch	0	0	0
24	Marino	Jugador23	23	183	No-Sin lesión	0	No-Sin lesión	0	No	0	1	0
25	Marino	Jugador24	26	180	ontractura de gemelos-Gemelos-PIERN	3	Contractura de sóleo-Sóleo-PIERNA ; F	3	Cubito radio	0	0	0
26	Marino	Jugador25	23	178	No-Sin lesión	0	Tendinitis rotuliana-Rodilla-PIERNA	1	No	0	0	0
27	Marino	Jugador26	23	185	o ; Rotura de cuádriceps-Cuádriceps-I	3	No-Sin lesión	0	No	0	0	0
28	Marino	Jugador27	27	185	No-Sin lesión	0	Esguince de tobillo-Tobillo-PIE	1	No	0	0	0
29	Marino	Jugador28	29	188	ingle-CADERA ; Hernia inguinal-ingle-	4	Esguince de tobillo-Tobillo-PIE	1	ibia perone/hernia	0	0	0
30	Marino	Jugador29	27	192	Rotura de menisco-Rodilla-PIERNA	1	res-Aductores-MUSLO ; Rotura de BF	2	Menisco int rod iz.	0	0	0
31	Marino	Jugador30	22	180	No-Sin lesión	0	No-Sin lesión	0	No	0	1	0
32	Marino	Jugador31	21	173	otibiales-Isquiosurales-MUSLO ; Rotu	5	IE ; Esguince de tobillo-Tobillo-PIE ; Es	3	No	0	0	0

Figura 2.1: Datos básicos de los jugadores

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Equipo	Jugador	V/V CALCÁNEO	DEBILIDAD FL. 1º DEDO	V/V RODILLA	PUNTILLAS IZD	PUNTILLAS D	PTTD (AMBOS)	TALÓN IZDA	TALÓN DCHA	D. UNIPODAL	D. BIPODAL	GLÚTEO MEDIO
2	Padre Anchieta	Jugador1	Valgo izquierdo	No	Neutro	12.5	12	13.5/14	14.5	14	No	No	No
3	Padre Anchieta	Jugador2	No	Derecho	Varo	10.5	12	12/12	12	12	Si	No	No
4	Padre Anchieta	Jugador3	No	Derecho	Neutro	11	11	11.5/11.5	10	10	No	No	No
5	Padre Anchieta	Jugador4	No	No	Neutro	13	13.5	13/14	13	13	No	Si	Izquierdo
6	Padre Anchieta	Jugador5	No	No	Varo	9.5	10	11/10	10	9	No	Si	No
7	Padre Anchieta	Jugador6	No	No	Valgo	10.5	11	14/15	10.5	11	No	Si	Ambos
8	Padre Anchieta	Jugador7	No	No	Varo	13.5	12	14.5/14	12.5	15.5	Si	No	No
9	Padre Anchieta	Jugador8	Valgo izquierdo	Derecho	Valgo	11.5	9.5	12.5/12	13	12	Si	No	Derecho
10	Padre Anchieta	Jugador9	No	No	Neutro	12	12.5	14/15	11	11.5	No	Si	Derecho
11	Padre Anchieta	Jugador10	No	Derecho	Neutro	13	14	13/15	10	9.5	Si	Si	Izquierdo
12	Padre Anchieta	Jugador11	No	No	Neutro	14.5	14	17/17	14.5	13	No	No	No
13	Padre Anchieta	Jugador12	Valgo ambos	No	Neutro	13.5	10	15/15	8.5	8.5	Si	No	No
14	Padre Anchieta	Jugador13	No	No	Varo	11	11.5	14/14	10	10	No	No	No
15	Padre Anchieta	Jugador14	Valgo izquierdo	No	Varo	10.5	9.5	10.5/11	11	11.5	Si	No	Derecho
16	Padre Anchieta	Jugador15	No	No	Neutro	9	9.5	9/10	8.5	7.5	No	Si	Izquierdo
17	Padre Anchieta	Jugador16	No	No	Neutro	12	13.5	14.5/15	11.5	11.5	Si	No	Izquierdo
18	Marino	Jugador17	No	No	Neutro	12	12	14.5/14	13	12	No	No	No
19	Marino	Jugador18	No	Izquierdo	Varo	13.5	12	15/14.5	13.5	13.5	Si	No	Ambos
20	Marino	Jugador19	No	No	Neutro	12.5	12.5	13.5/13	12	12	No	No	No
21	Marino	Jugador20	Valgo ambos	No	Neutro	13	13.5	14.5/14.5	11.5	12	No	No	Derecho
22	Marino	Jugador21	No	Derecho	Neutro	13	13	15/15	10.5	10	No	No	No
23	Marino	Jugador22	No	Ambos	Varo	7	10	15/16	13	13	Si	Si	No
24	Marino	Jugador23	Valgo ambos	Derecho	Neutro	13	13	16/15	11	10	No	Si	No
25	Marino	Jugador24	No	No	Neutro	13.5	13.5	16/16	13	15	No	No	No
26	Marino	Jugador25	Valgo ambos	No	Varo	8	8	14/14	13	14	No	No	Izquierdo
27	Marino	Jugador26	No	No	Varo	12	12	14/14	12	13	No	No	No
28	Marino	Jugador27	Valgo ambos	No	Varo	13.5	14.5	14/14	13	12	Si	No	No
29	Marino	Jugador28	No	No	Varo	13.5	14.5	14.5/15.5	13	13	Si	Si	Derecho

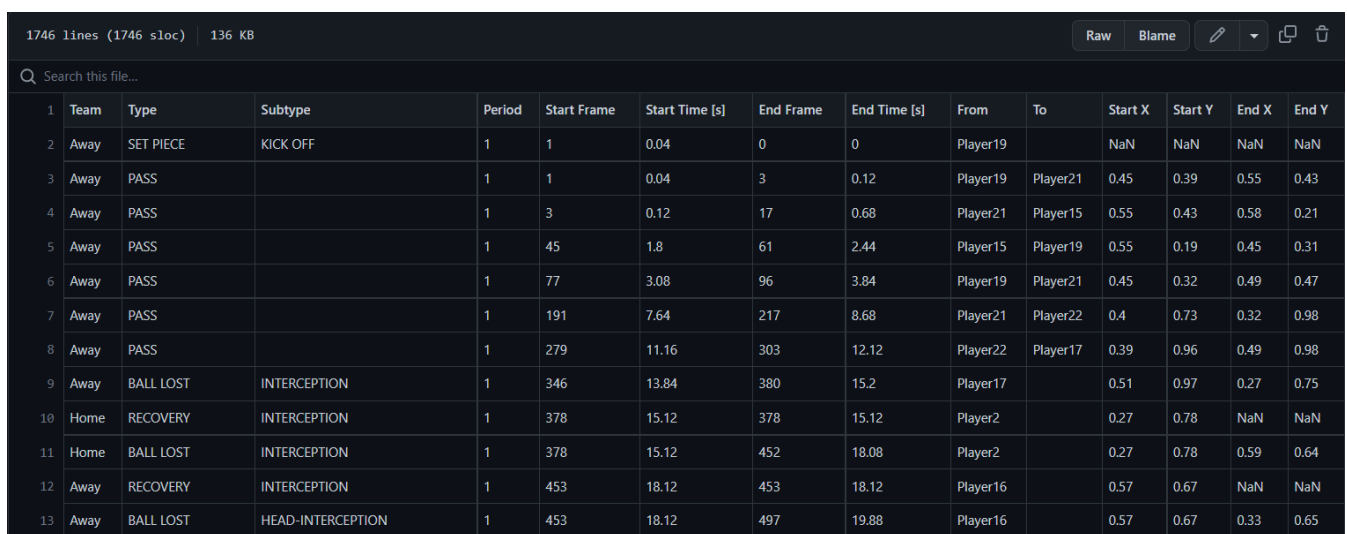
Figura 2.2: Factores de riesgo de los jugadores



## 2.3.2. Data-tracking

El segundo estudio trata sobre data-tracking, para los cuales necesitamos unos datos muy específicos como la posición de los jugadores y el balón en todo el momento del partido, los eventos ocurridos, los tipos de eventos, momento inicial y final de las jugadas, etc. Por suerte, Metrica-Sports, un software dedicado al data-tracking de fútbol, nos ofrece estos datos mencionados a modo de ejemplo con el que poder realizar el estudio. Los datos están divididos en 3 partidos, donde cada uno contiene 3 ficheros diferentes:

- **Eventos.** Fichero CSV cuya línea representa un evento, un pase, un disparo, un fuera de juego, falta, etc. (Figura 2.3)
- **Posiciones.** 2 ficheros CSV, uno para el equipo local y otro para el equipo visitante, estos, contienen la posición de cada jugador así como del balón cada 1/4 de segundo.



	Team	Type	Subtype	Period	Start Frame	Start Time [s]	End Frame	End Time [s]	From	To	Start X	Start Y	End X	End Y
1	Away	SET PIECE	KICK OFF	1	1	0.04	0	0	Player19		NaN	NaN	NaN	NaN
2	Away	PASS		1	1	0.04	3	0.12	Player19	Player21	0.45	0.39	0.55	0.43
3	Away	PASS		1	3	0.12	17	0.68	Player21	Player15	0.55	0.43	0.58	0.21
4	Away	PASS		1	45	1.8	61	2.44	Player15	Player19	0.55	0.19	0.45	0.31
5	Away	PASS		1	77	3.08	96	3.84	Player19	Player21	0.45	0.32	0.49	0.47
6	Away	PASS		1	191	7.64	217	8.68	Player21	Player22	0.4	0.73	0.32	0.98
7	Away	PASS		1	279	11.16	303	12.12	Player22	Player17	0.39	0.96	0.49	0.98
8	Away	BALL LOST	INTERCEPTION	1	346	13.84	380	15.2	Player17		0.51	0.97	0.27	0.75
9	Home	RECOVERY	INTERCEPTION	1	378	15.12	378	15.12	Player2		0.27	0.78	NaN	NaN
10	Home	BALL LOST	INTERCEPTION	1	378	15.12	452	18.08	Player2		0.27	0.78	0.59	0.64
11	Away	RECOVERY	INTERCEPTION	1	453	18.12	453	18.12	Player16		0.57	0.67	NaN	NaN
12	Away	BALL LOST	HEAD-INTERCEPTION	1	453	18.12	497	19.88	Player16		0.57	0.67	0.33	0.65

Figura 2.3: Tabla de los eventos del partido

## 2.4. Metodología

Una vez obtenidos los datos, tenemos que realizar una serie de procesos, los cuales explicaremos más en detalle a lo largo del trabajo:

1. Obtención de los datos
2. Limpieza de los datos (Modificación o eliminación de columnas)
3. Análisis de los datos (Creación de nuevos datos)
4. Visualización de los datos a través de herramientas gráficas
5. Obtención de información a partir de los nuevos datos

# Capítulo 3

## Objetivos

En este capítulo explicaremos los diferentes objetivos generales así como los específicos de este estudio.

### 3.1. Objetivo general

El objetivo de este TFG es mostrar como a partir de una serie de datos, podemos procesarlos, modificarlos y hacer un análisis exhaustivo de los mismos sobre diferentes ámbitos de la practica del fútbol. Estos son las diferentes lesiones que pueden sufrir los jugadores, el rendimiento de los equipos cuando recorren grandes distancias y el análisis de las posiciones y eventos de un partido de fútbol.

El fin es poder hacer ese análisis completo para poder obtener un conocimiento con el que poder apoyar la toma de decisiones de los diferentes equipos técnicos de los equipos de fútbol con el que podrán desarrollar sus habilidades y resultados.

### 3.2. Objetivos específicos

#### 3.2.1. Estudio de lesiones

Este estudio tiene como fin el análisis de las lesiones de diferentes jugadores de varios equipos de la categoría amateur de Tenerife, para poder saber cuales son las lesiones más frecuentes, los factores de riesgos que más afectan a las diferentes grupos musculares, lesiones, etc, con lo cual poder aportar a su prevención.

#### 3.2.2. Estudio de data-tracking

Por último tenemos este estudio, cuyo objetivo es hacer un análisis del rendimiento individual de los jugadores en un partido de fútbol, poder saber todos los movimientos que realizo, pases acertados, disparos, etc. Así como el análisis del juego colectivo de ambos equipos, por último haremos un análisis de las probabilidades de éxito que tiene un pase con el fin de mejorar el posicionamiento defensivo y las disposición técnica de los jugadores.

# Capítulo 4

## Desarrollo

En este capítulo explicaremos en detalle el proceso que hemos seguido en los casos de estudio así como los resultados obtenidos.

### 4.1. Entorno de desarrollo

Para poder desarrollar el trabajo con facilidad hemos hecho uso de un entorno Linux dentro de Windows 10, conocido como WSL (Windows Subsystem for Linux). Como editor de código utilizaremos Visual Studio Code, a través del cual tenemos acceso al subsistema Linux mencionado.

Solo quedaría instalar Python, las librerías o paquetes que necesitamos (Pandas, Matplotlib...) y las extensiones para Visual Studio Code. Por otro lado, también crearemos un entorno similar en Google Colab, donde podemos almacenar el código y ejecutarlo en la nube. Para mantener el código en local actualizado con el de Google Colab utilizábamos las herramientas de git, subíamos el código desde local y en Colab hacíamos un pull para recoger los cambios.

### 4.2. Caso 1: Lesiones

Los datos del primer caso están formados por dos tablas, una de la información básica del jugador y otra de los factores de riesgo de cada uno de ellos.

- Información básica (Columnas):

Equipo	Jugador	Edad	Lesiones Previas	Total L.P
Lesiones Actuales	Total L.A	Operaciones	Diabetes	Corrección Ort.
Escoliosis	Pie Dominante	Posición	Preparación	Calzado

Tabla 4.1: Tabla de la información básica del jugador

- Factores de riesgo (Columnas):

Equipo	Jugador	V/V Calcáneo	Debilidad FL. 1º Dedo
D. Unipodal	D. Bipodal	Glúteo Medio	Anteversión-Retroversión
Sentadilla	Sprint	V/V Rodilla	Overhead Squat

Tabla 4.2: Tabla de los distintos factores de riesgo

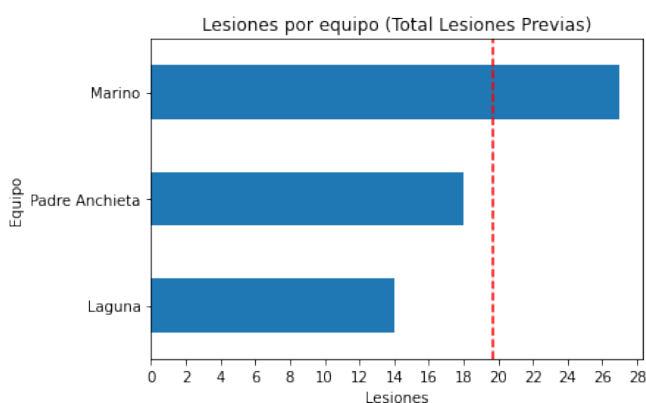


Figura 4.1: Número de lesiones por equipo (Previas)

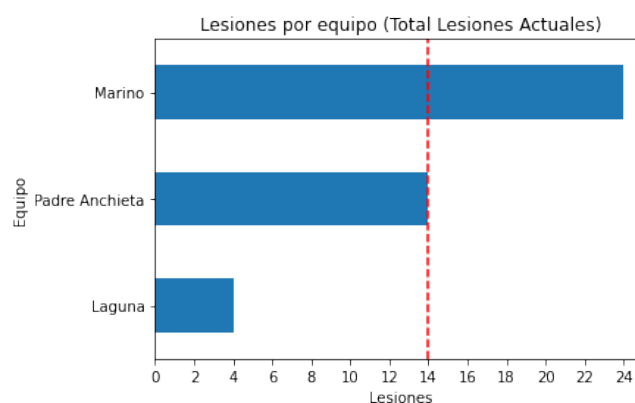


Figura 4.2: Número de lesiones por equipo (Actuales)

### 4.2.1. Visualizaciones Básicas

Estas visualizaciones nos dan una idea inicial de los datos, no es necesaria ningún tipo de modificación:

- Número de lesiones por equipo (Figuras 4.1 y 4.2)
- Número de lesiones por equipo agrupadas por posición de los jugadores (Figura 4.3)

Registramos tanto las lesiones previas como las ocurridas a lo largo de esta temporada 2021-2022, contabilizando un total de 101 lesiones (59 en temporadas previas y 42 actuales). En cuanto al total de lesiones (tanto esta temporada como en anteriores), los jugadores del C.D. Marino son los que más presentaron, seguidos de los jugadores del Padre Anchieta y el C.D. Laguna. La media de lesiones este año es de 14 (Marino 24 lesiones, Anchieta 14 lesiones y Laguna 4 lesiones).

Si comparamos las lesiones previas (Figura 4.1) con las actuales (Figura 4.2) podemos observar que el equipo Laguna ha tenido una mejora con respecto a las lesiones, bajando en un 70%.

En el cómputo global de lesiones, los delanteros son los que más lesiones sufren (34), seguido de los defensas (33) y los medio centros (29). Un mayor número de lesiones previas se registró en los delanteros, mientras que esta temporada el grupo más afectado fue el de los defensores.

### 4.2.2. Modificación de los datos

Originalmente la columna Lesiones Previas venía estructurada como un vector de lesiones, separada por ';', cada lesión, descrita por su región y grupo muscular o articulación afectada. Por ejemplo: Esguince de tobillo-Tobillo-PIE

Sabiendo esto pudimos operar sobre esta columna, haciendo un *split* de cada una de las lesiones de los jugadores, obteniendo una nueva columna con las lesiones en formato *array* de Python, con la ayuda del método *explode* de Pandas obteníamos nuevas filas, una fila adicional por cada lesión del jugador. (Figura 4.9)

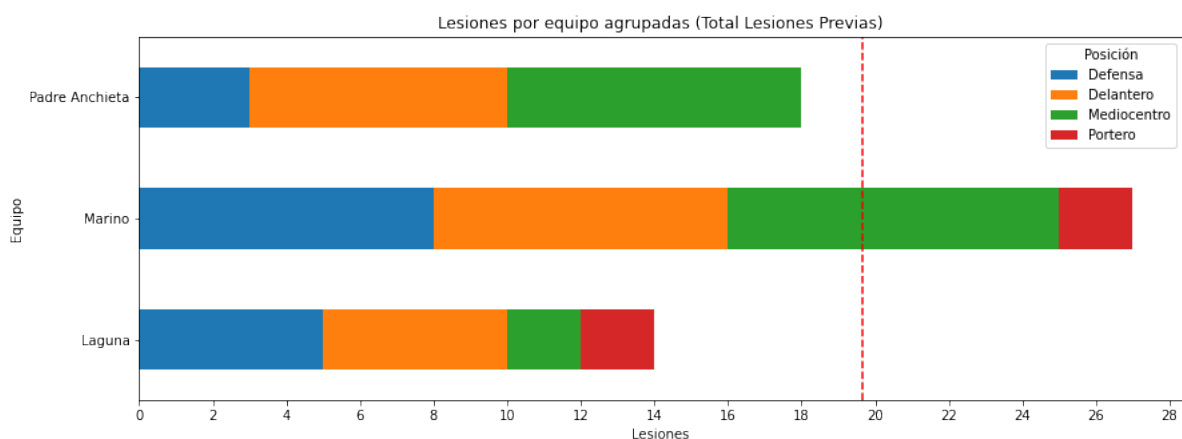


Figura 4.3: Número de lesiones por equipo agrupadas por posición de los jugadores (Previas)

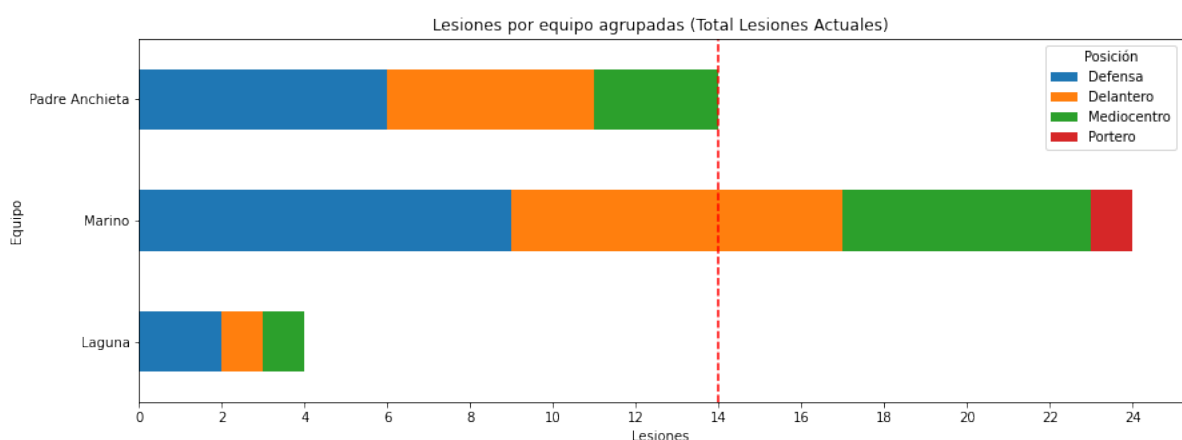


Figura 4.4: Número de lesiones por equipo agrupadas por posición de los jugadores (Actuales)

Una vez modificados los datos pudimos obtener las gráficas de cuales eran las zonas más afectadas en las lesiones. En el total registrado, la región más afectada fue el muslo y el pie con el mismo número de incidencias (30 lesiones), seguido de la pierna (26). Este año, la región más afectada fue el muslo (17), seguida por el pie (14) y la pierna (8) (Figura 4.6). En años anteriores, se registraron más lesiones en la pierna (18), seguido del pie (16) y el muslo (13) (Figura 4.5). El registro de lesiones acumuladas señala al tobillo como la parte más afectada (28 incidencias), seguido de lesiones de rodilla e isquiosurales a partes iguales (16), cuádriceps (9), tríceps sural (6) y aductores (5). Este año, las mayores incidencias registradas fueron en tobillo (12) e isquiosurales (10) (Figura 4.8). En el apéndice A se incluye los algoritmos utilizados para poder crear los gráficos.

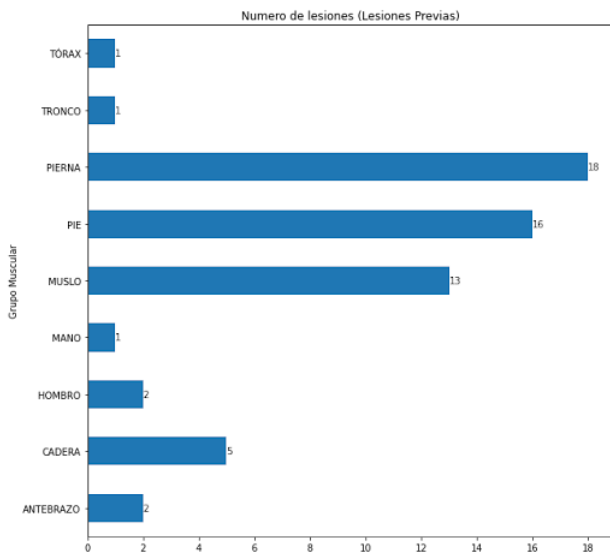


Figura 4.5: Total lesiones por zona (Previas)

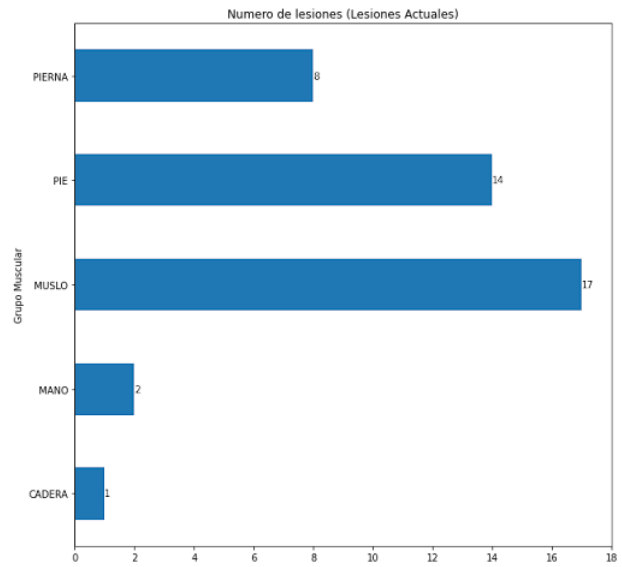


Figura 4.6: Total lesiones por zona (Actuales)

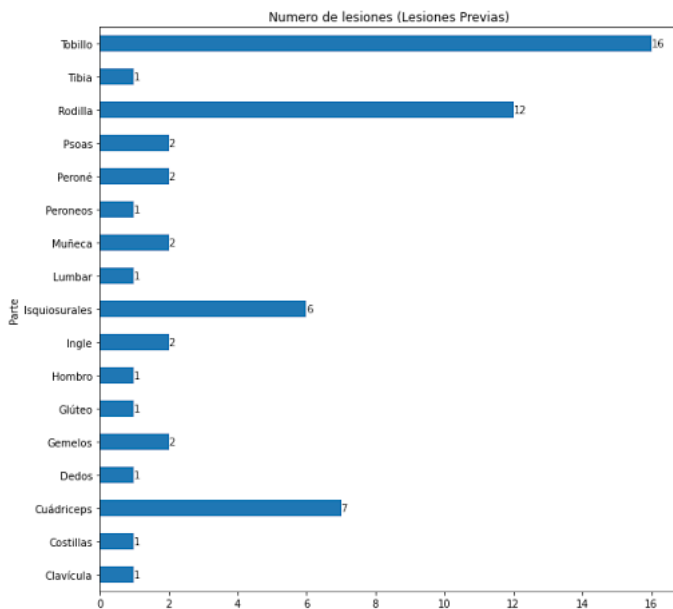


Figura 4.7: Total lesiones por parte (Previas)

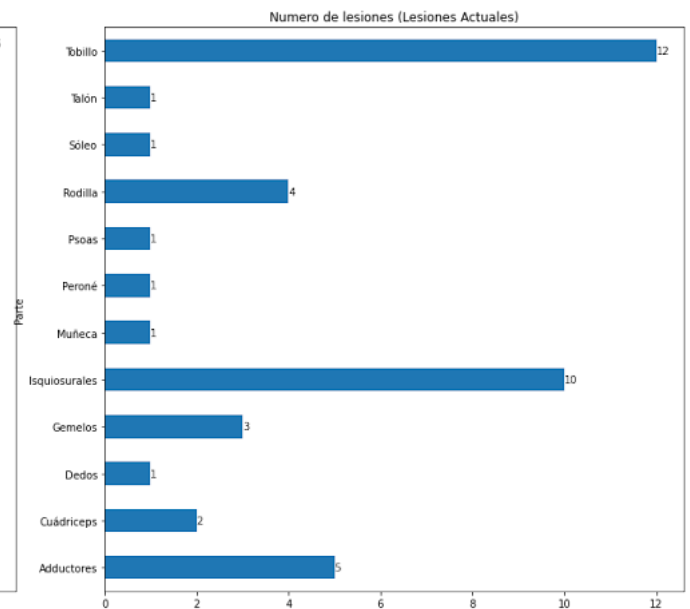


Figura 4.8: Total lesiones por parte (Actuales)

	Jugador	Equipo	Edad	Altura	Vector Lesiones	Lesion	Parte	Grupo Muscular
0	Luis Pastor	Padre An...	22	181	Fractura de clavícula-Clavícula-HOMBRO	Fractura de clavícula	Clavícula	HOMBRO
1	Luis Pastor	Padre An...	22	181	Esguince de tobillo-Tobillo-PIE	Esguince de tobillo	Tobillo	PIE
2	Luis Pastor	Padre An...	22	181	Esguince de tobillo-Tobillo-PIE	Esguince de tobillo	Tobillo	PIE
3	Luis Pastor	Padre An...	22	181	Rotura de recto anterior-Cuádriceps-M...	Rotura de recto anterior	Cuádriceps	MUSLO
4	Eduardo	Padre An...	22	185	Molestia de psoas-Psoas-CADERA	Molestia de psoas	Psoas	CADERA
5	Alberto B...	Padre An...	20	180	Luxación de hombro-Hombro-HOMBRO	Luxación de hombro	Hombro	HOMBRO
6	Omar	Padre An...	26	175	No-Sin lesión	No	Sin lesión	Sin lesión
7	Keyber	Padre An...	21	170	Esguince de tobillo-Tobillo-PIE	Esguince de tobillo	Tobillo	PIE

Figura 4.9: Modificación de la tabla a partir del método explode

### 4.2.3. Factores de Riesgo

Se denomina factor de riesgo a cualquier característica o circunstancia detectable de una persona que se sabe asociada con un aumento en la probabilidad de padecer, desarrollar o estar especialmente expuesto a una enfermedad. Para este caso estudiaremos los factores de riesgo que pueden afectar en la práctica del fútbol.

- Varo/Valgo calcáneo (**Desviación del talón**): 35 individuos sin V/V calcáneo y 16 con V/V calcáneo (8 valgo ambos; 3 valgo izquierdo, 2 varo izquierdo, 1 varo ambos, 1 varo derecho, 1 varo izquierdo) (Figura 4.10).
- Debilidad del flexor del primer dedo (**Musculo flexor del primer dedo del pie**): 35 individuos sin debilidad, y 16 con debilidad (11 derecho, 4 ambos, 1 izquierdo) (Figura 4.11).
- Varo/Valgo de rodilla (**Desviación de la rodilla**): 25 individuos con varo de rodilla, 23 con posición neutra y 3 con valgo de rodilla (Figura 4.12).
- Dismetría unipodal (**Activación del tibial posterior**): 28 sin disimetría unipodal y 23 con disimetría unipodal.
- Dismetría bipodal: 36 individuos sin disimetría bipodal y 15 con disimetría bipodal.
- Glúteo medio (**Estabilizador de la cadera**): 27 individuos sin debilidad de glúteo medio y 24 con debilidad (10 glúteo derecho, 9 glúteo izquierdo, 5 ambos glúteos) (Figura 4.13).
- Anteversión/Retroversión: 35 individuos con pelvis neutra, a 15 con anteversión y 1 con retroversión pélvica.
- Sprint: 31 jugadores con una técnica de carrera sin rotación externa de cadera y a 20 jugadores con una rotación externa de cadera.

Tras la obtención de las lesiones totales y los datos sobre los factores de riesgo de cada individuo, realizamos una asociación de ambos datos para observar cómo se comportan uno en relación al otro, y observar las posibles tendencias que puedan surgir de dicho emparejamiento. Se emparejó cada factor de riesgo individualmente con las lesiones totales, dando los siguientes resultados:

- Según Varo/Valgo calcáneo: Existe una alta incidencia lesional en tobillo ligada a la presencia de alguna de las posibilidades de este factor de riesgo (ya sea varo o valgo unilateral o bilateral), representando un 50 % de las lesiones totales de tobillo (esguinces) (Figura 4.14).

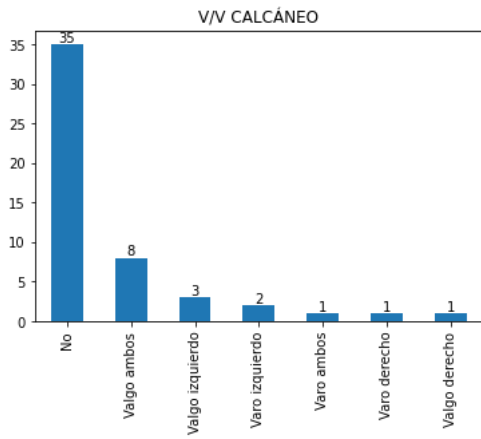


Figura 4.10: Jugadores con V/V Calcáneo

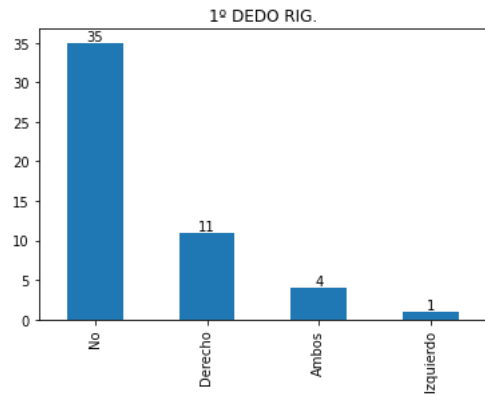


Figura 4.11: Jugadores con 1º dedo rígido

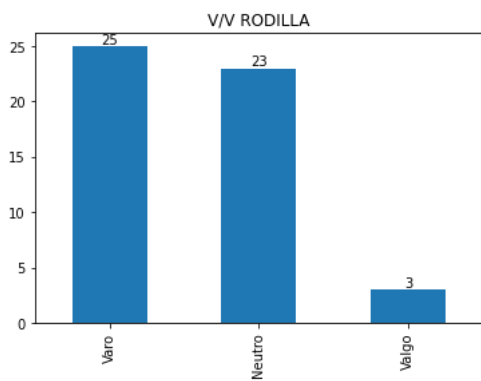


Figura 4.12: Jugadores con V/V Rodilla

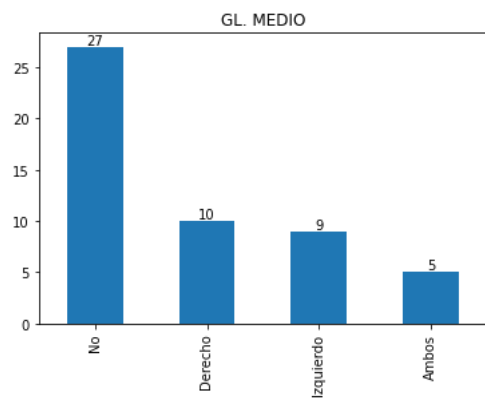


Figura 4.13: Jugadores con Glúteo Medio



- Según debilidad del flexor del primer dedo: Se observó una moderada incidencia lesional en el grupo de tobillo (28.57 %) asociada a este factor de riesgo (Figura 4.17).
- Según Varo/Valgo de rodilla: Se ha observado en alta medida la presencia de varo de rodilla en lesiones de rodilla (62.5 %), de isquiosurales (62.5 %) y de tobillo (75 %) (Figura 4.15).
- Según debilidad de glúteo medio: En cuanto a la presencia de debilidad del glúteo medio (tanto unilateral como bilateral) y la tasa de lesiones, la mayor relación significativa que existe es con las lesiones de rodilla (62.5 %), y cuádriceps (44.44 %) (Figura 4.16).

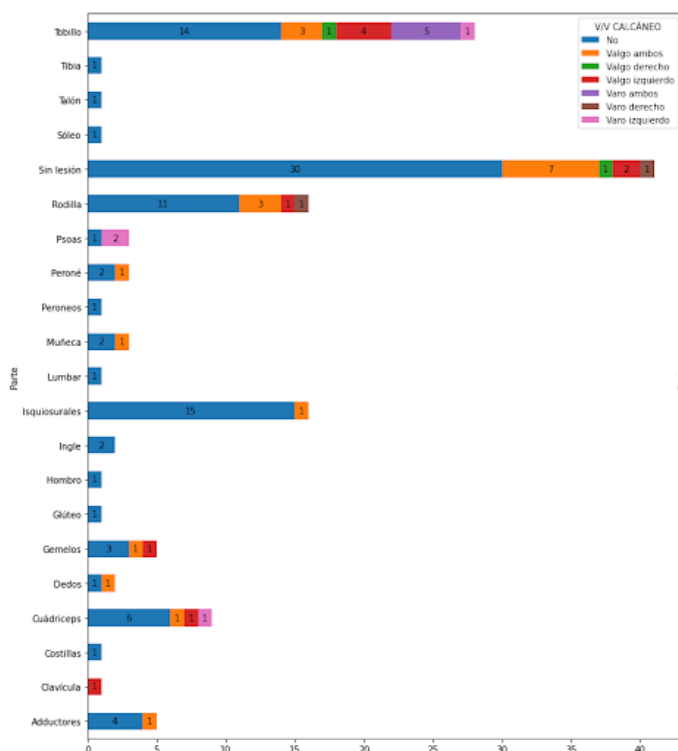


Figura 4.14: Lesiones con V/V Calcáneo

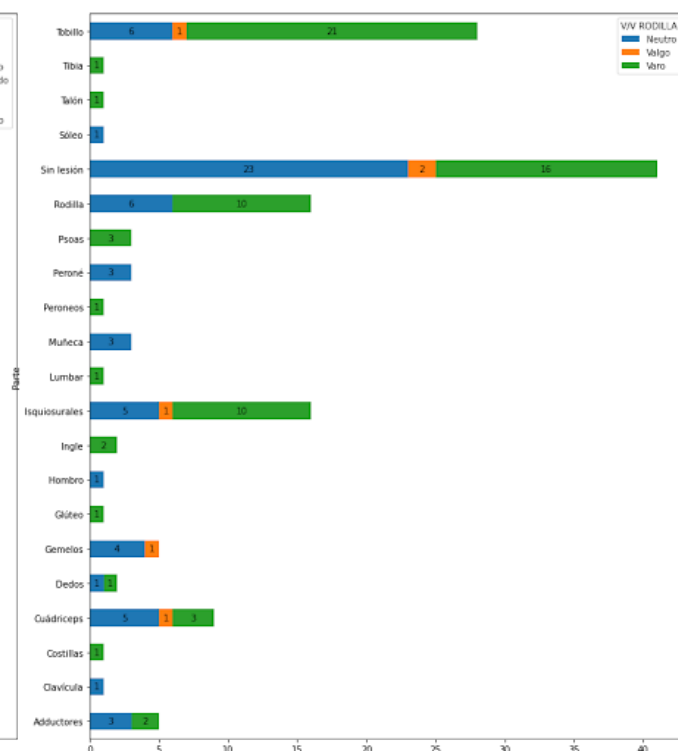


Figura 4.15: Lesiones con V/V Rodilla

### 4.3. Caso 2: Data-tracking

Como mencionamos anteriormente este caso es el análisis de las estadísticas y posicionamiento táctico de un partido de fútbol. El partido se divide en tres ficheros diferentes:

- RawEventsData.csv. Contiene los eventos de todo el partido, pases, disparos, fuera de juegos, saques de puerta, goles, etc. También incluye las coordenadas x e y del inicio y del final de cada uno de estos eventos.
- RawTrackingData\_Away\_Team.csv. Contiene la coordenadas x e y de cada jugador del equipo visitante, las mediciones están hechas con una frecuencia de 25Hz, es decir, 4 veces por segundo.
- RawTrackingData\_Home\_Team.csv. Contiene la coordenadas x e y de cada jugador del equipo local, las mediciones están hechas con una frecuencia de 25Hz, es decir, 4 veces por segundo. Además contiene la posición del balón.

#### 4.3.1. Modificación inicial de los datos

La forma en la que está estructurados los ficheros de las posiciones de los jugadores no es la más adecuada (Figura 4.18), por lo tanto es necesario hacer una pequeña modificación para que sean más legibles y fáciles de analizar.

La tabla quedaría con varias columnas adicionales, dos para cada jugador, la coordenada x (away\_Player\_x) y la coordenada y (away\_Player\_y) (Figura 4.19). Adicionalmente, para ayudar a legibilidad de los datos hemos añadidos nombre reales a los jugadores, simulando un partido entre la selección francesa masculina contra la selección española masculina, visitante y local, respectivamente.

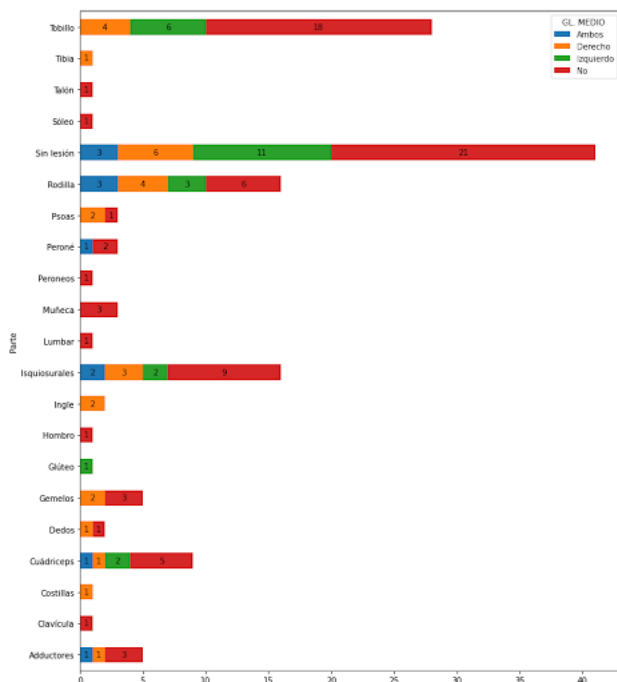


Figura 4.16: Lesiones con Gl. Medio

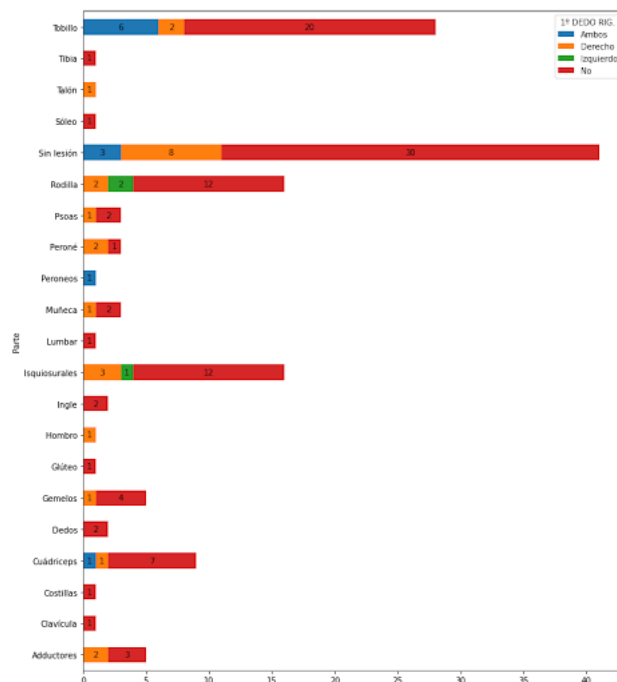


Figura 4.17: Lesiones con 1º dedo rígido

```

1 , , , Away , Away , Away , Away , Away , Away , Away , Away , Away , Away , Away , Away , , ,
2 , , , 25 , 15 , 16 , 17 , 18 , 19 , 20 , 21 , 22 , 23 , 24 , 26 , , ,
3 Period,Frame,Time [s],Player25,Player15,Player16,Player17,Player18,,Player19,,Player20,,Player21,,Player22,,
4 1,1,0.04,0.01218,0.51763,0.32891,0.81846,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.3981
5 1,2,0.08,0.01218,0.51763,0.32891,0.81846,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.3981
6 1,3,0.12,0.01218,0.51763,0.32891,0.81846,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.3981
7 1,4,0.16,0.01218,0.51763,0.32836,0.81847,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.3981
8 1,5,0.2,0.01218,0.51763,0.32783,0.81847,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.39812
9 1,6,0.24,0.01218,0.51763,0.3273,0.81847,0.30734,0.55182,0.32235,0.38954,0.36505,0.31049,0.49564,0.91364,0.39812
10 1,7,0.28,0.01064,0.51507,0.32678,0.81847,0.30816,0.55982,0.32379,0.41331,0.36505,0.31049,0.49564,0.91364,0.39995
11 1,8,0.32,0.00909,0.5125,0.32626,0.81848,0.30837,0.56142,0.32395,0.41814,0.36505,0.31049,0.49564,0.91364,0.39986
12 1,9,0.36,0.00754,0.50991,0.32573,0.81848,0.30844,0.56203,0.32387,0.42022,0.36505,0.31049,0.49564,0.91364,0.40000
13 1,10,0.4,0.00598,0.50732,0.32522,0.81848,0.30849,0.56249,0.32376,0.42134,0.36505,0.31049,0.49564,0.91364,0.40000

```

Figura 4.18: Formato inicial de los datos de posiciones de los jugadores

Frame	Period	Time [s]	away_Lloris1_x	away_Lloris1_y	away_Pavard2_x	away_Pavard2_y	
0	1	0.04	-51.71	-1.2	-18.14	-21.66	
1	2	1	0.08	-51.71	-1.2	-18.14	-21.66
2	3	1	0.12	-51.71	-1.2	-18.14	-21.66
3	4	1	0.16	-51.71	-1.2	-18.19	-21.66
4	5	1	0.2	-51.71	-1.2	-18.25	-21.66
5	6	1	0.24	-51.71	-1.2	-18.31	-21.66
6	7	1	0.28	-51.87	-1.02	-18.36	-21.66
7	8	1	0.32	-52.04	-0.85	-18.42	-21.66
8	9	1	0.36	-52.2	-0.67	-18.47	-21.66
9	10	1	0.4	-52.37	-0.5	-18.53	-21.66

Figura 4.19: Formato de los datos de las posiciones modificados

Otra de las modificaciones a estos datos fue cambiar el formato de la coordenadas, tanto de los eventos como la de las posiciones, ya que estos estaban en un rango de valores de 0 a 1, las pasamos a formato métrico, poniendo como medidas del campo 106 x 68 metros, siendo el punto central del campo el (0,0). En la figura 4.19 podemos ver las coordenadas del jugador Lloris, ubicado en la -51.71, -1.2 en el inicio del partido, por lo

tanto, deducimos que se encuentra cerca de la portería izquierda.

### 4.3.2. Visualizaciones básicas

Una vez modificada las unidades de las coordenadas podemos mostrarlas en un gráfico que tuviera las mismas dimensiones que el campo (106x68). Para esto lo ideal sería crear dicho campo, con las porterías, círculo central, bandas, etc.

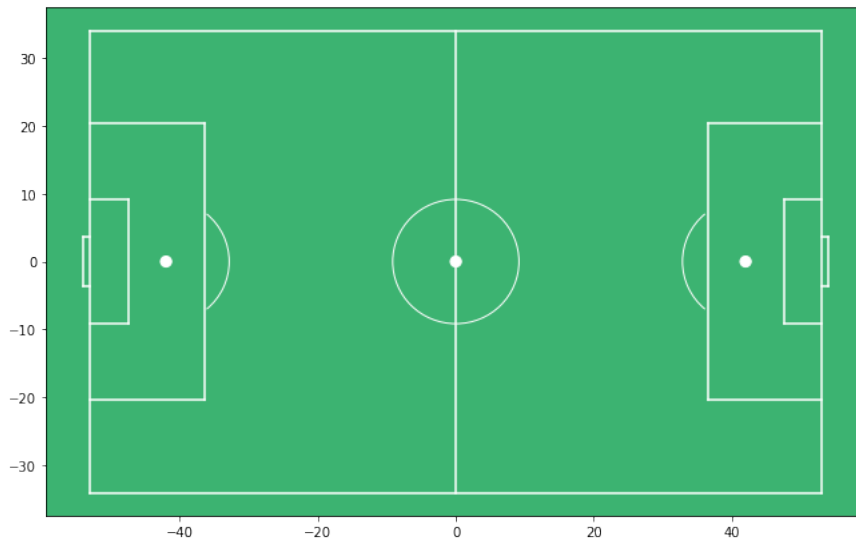


Figura 4.20: Gráfico del terreno de juego

Podemos mostrar un evento dibujando una línea sobre el terreno de juego, por ejemplo los puntos inicial y final de un pase:

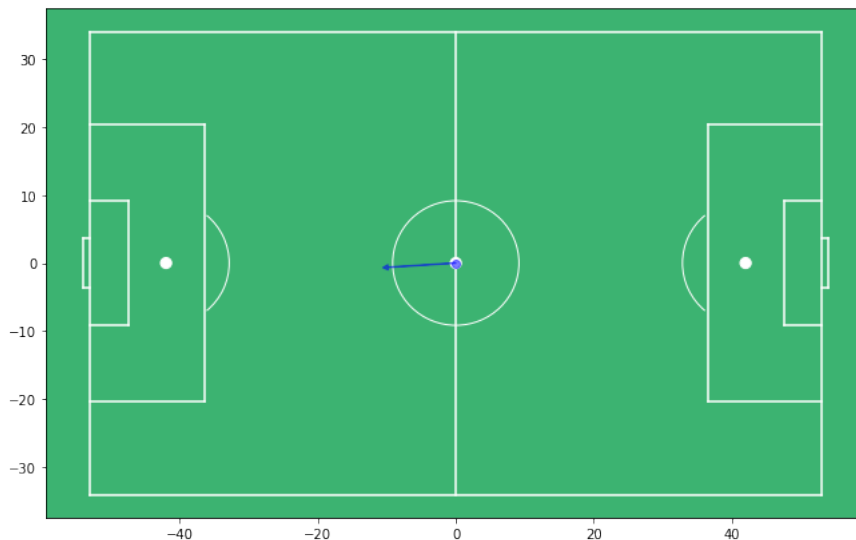


Figura 4.21: Primer evento del partido

También podemos imprimir la posición de cada jugador en un frame concreto:

Gracias a esta imagen podemos sacar un serie de conclusiones, viendo los datos en crudo no obtendríamos ningún tipo de información al respecto de la disposición técnica de los equipos, pero gracias al método de dibujar los jugadores podemos observar que el equipo azul (Francia) empieza en el lado izquierdo del campo, por ende, el equipo rojo,

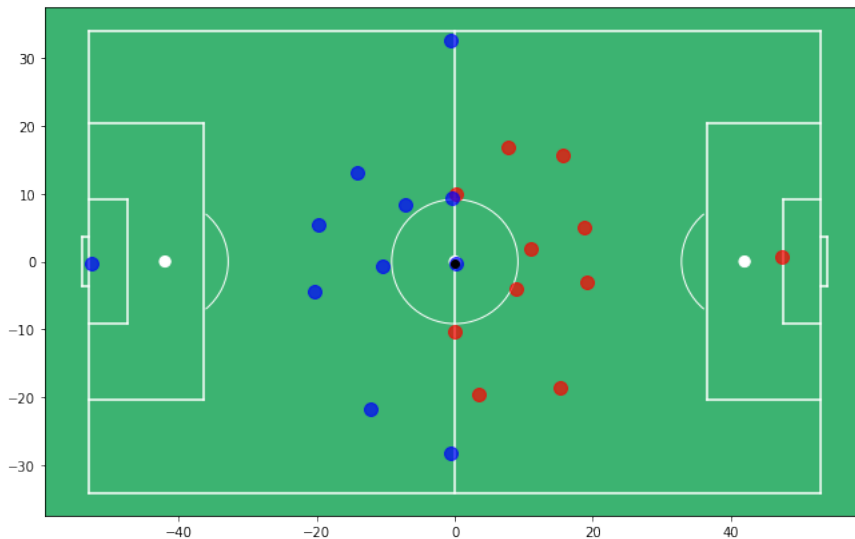


Figura 4.22: Posición inicial de los jugadores y el balón

por el lado derecho. Ambos juegan de inicio con el mismo sistema táctico 4-4-2, dos líneas de 4 jugadores y una línea de 2 jugadores.

Podemos incluir los nombres de los jugadores para tener una mejor idea del sistema táctico:

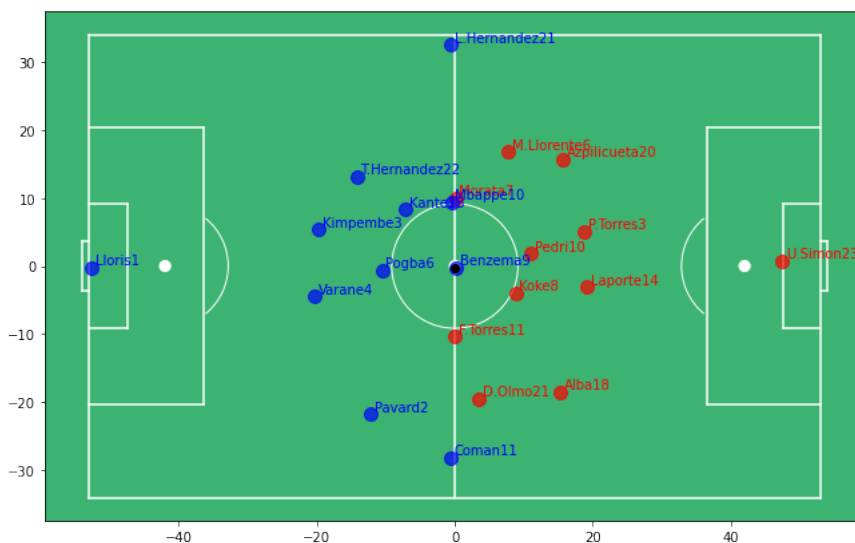


Figura 4.23: Posición inicial con nombres

### 4.3.3. Calculando velocidad

Sabemos que los datos contienen información sobre las posiciones de los jugadores en un diferentes instantes de tiempo, por lo tanto, con esto ya podemos sacar la velocidad de los jugadores en cada instante. Calcularemos la velocidad en la dirección x e y como la velocidad total. Para ejecutar esto solo tenemos que dividir la diferencia de posición en x e y de cada una de las filas entre la diferencia de tiempo de las mismas.

Podemos añadir en el método que dibujaba los jugadores una opción para que nos muestre la velocidad de cada jugador mediante una flecha, cuanto más larga la flecha,

mas velocidad. En el apéndice B se incluye el algoritmo para hacer el cálculo de las velocidades.

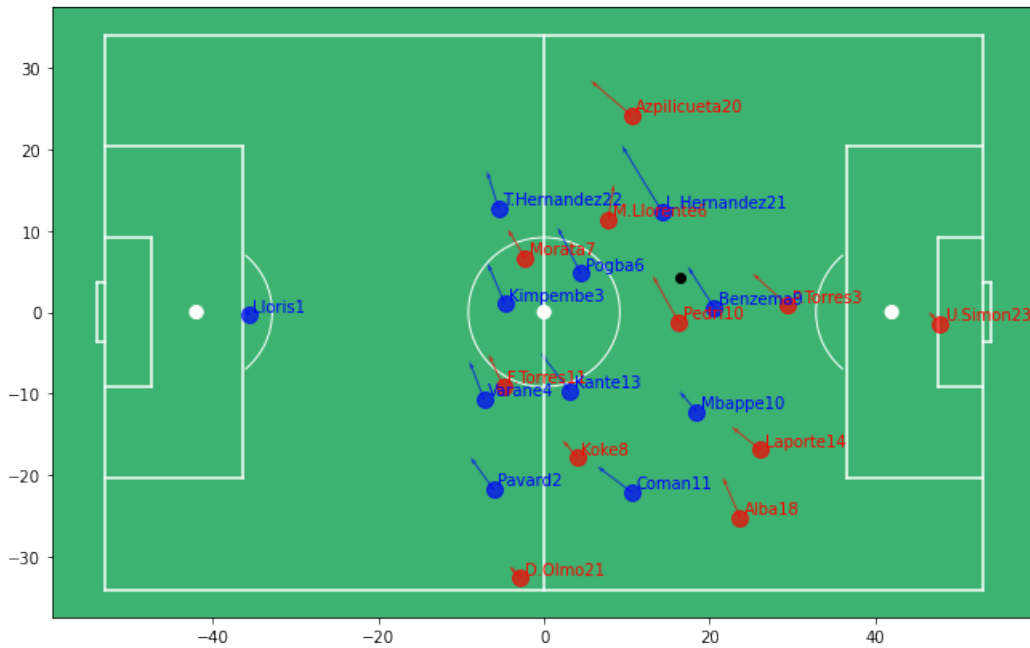


Figura 4.24: Velocidad de los jugadores

Ahora con los datos de velocidad podemos sacar algunas estadísticas sobre distancia recorrida por los jugadores:

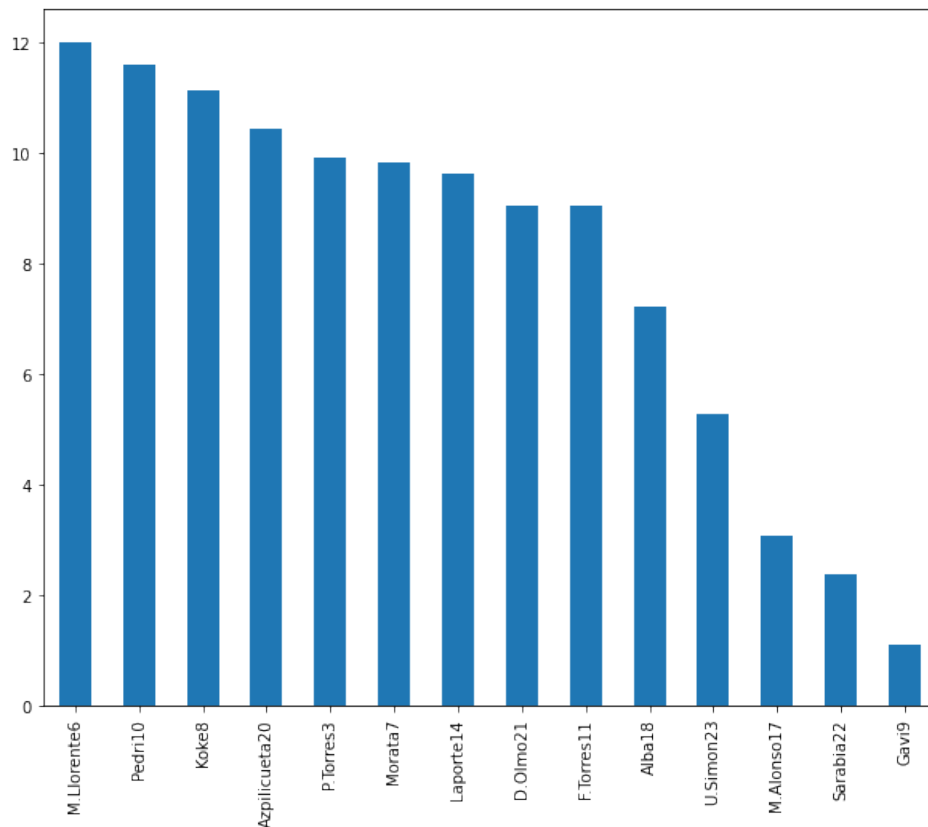


Figura 4.25: Distancia recorrida por cada jugador del equipo local (Km)

Además, podemos agrupar esta distancia que recorrió cada jugador clasificadas según

a la velocidad que lo hizo, cabe destacar que la velocidad esta medida en m/s, podemos clasificarla en 4 puntos:

- Caminando (Walking): velocidad menor a 2 m/s.
- Trotando (Jogging): velocidad entre 2 y 4 m/s.
- Corriendo (Running): velocidad entre 4 y 7 m/s.
- Esprintando (Sprinting): velocidad superior a 7 m/s.

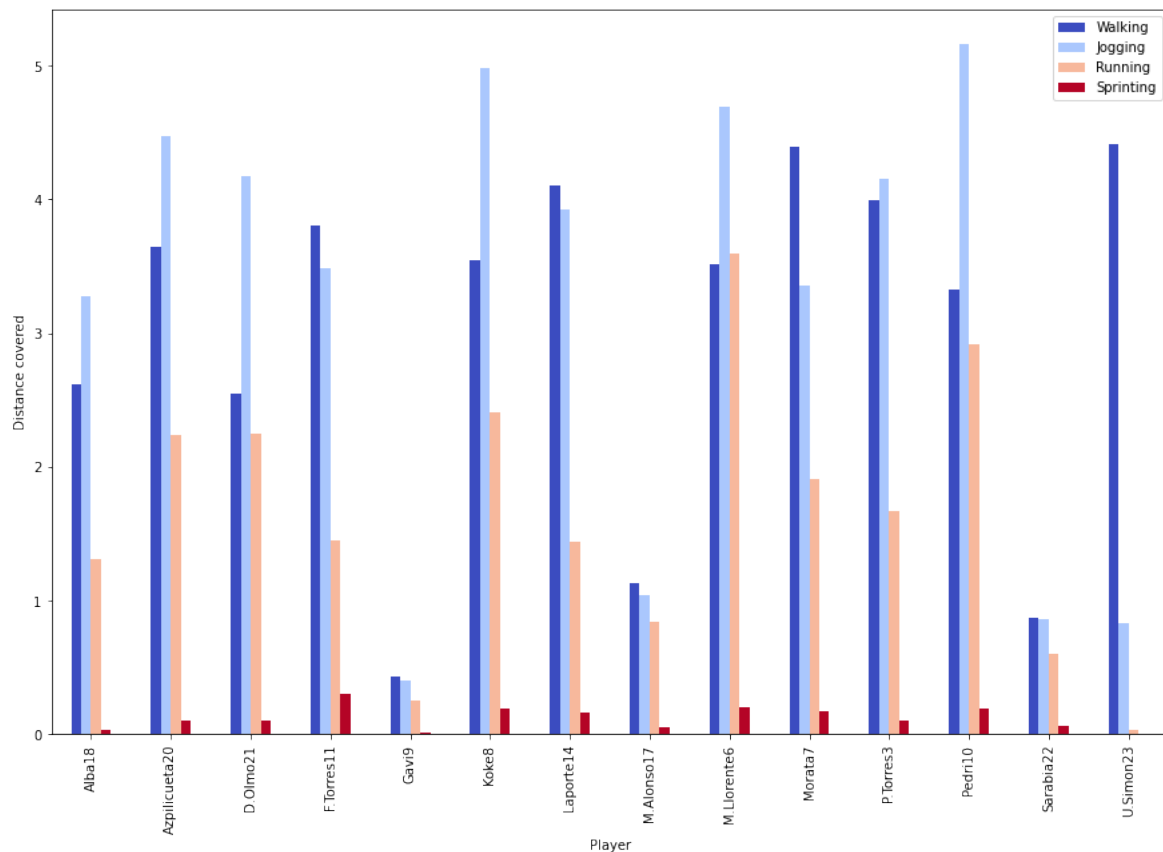


Figura 4.26: Distancia recorrida por jugador clasificada por velocidad

Podemos observar que la mayoría del tiempo los jugadores no están corriendo a toda velocidad, es más, la distancia que recorren esprintando es mínima. Observamos que los jugadores que mas distancia recorrieron fueron a alta velocidad fueron los medio centros Koke y Pedri, junto a los jugadores pegados a la banda, M. Llorente y Azpilicueta, por lo tanto podemos afirmar que la mayor actividad del juego ofensivo y defensivo fue por la banda derecha del conjunto local.

#### 4.3.4. Estadísticas de equipo

Gracias a la tabla de los eventos podemos sacar las estadísticas más básicas que nos ofrecen siempre que vemos un partido de fútbol, como los pases, posesión, etc.

La posesión se puede calcular de varias maneras, mediante el método Reloj de Ajedrez, que implementa un sistema de botones para contar el tiempo que un equipo mantiene

la posición del balón. Luego, existe el método algorítmico, el cuál aplicaremos en este estudio.

$$P_a = \frac{Pases_a}{Pases_a + Pases_b}$$

Para hallar la posesión de balón del equipo a es tan sencillo como dividir los pases acertados del equipo a entre la suma de pases acertado del equipo a y b. Una vez calculada la posesión por ambas partes podemos mostrarla con un diagrama de sectores:

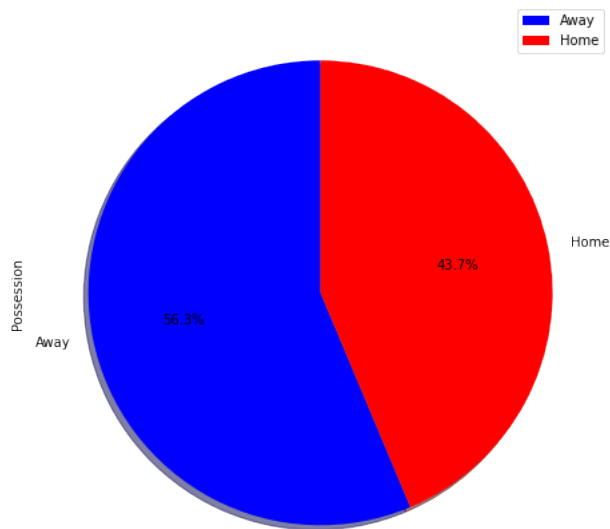


Figura 4.27: Posesión de balón

Estadísticas de los eventos del encuentro:

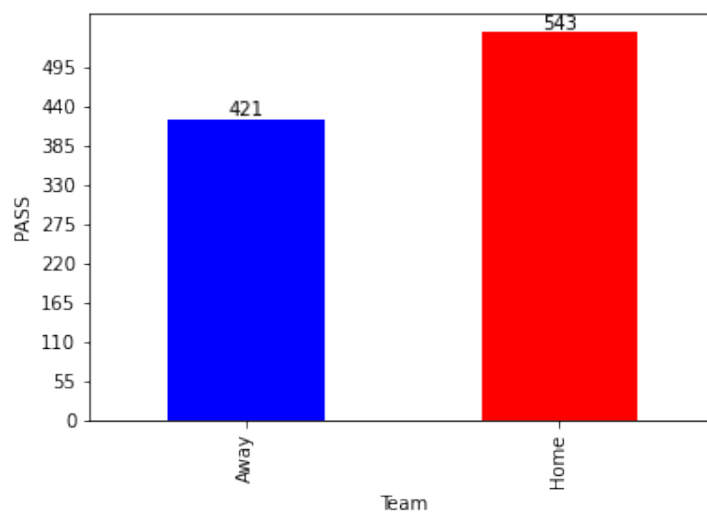


Figura 4.28: Pases acertados



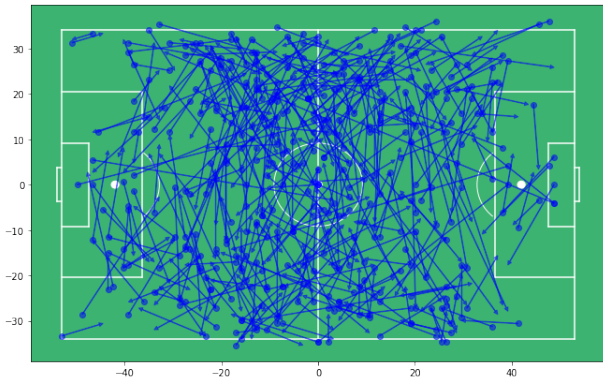


Figura 4.29: Pases completados del equipo visitante

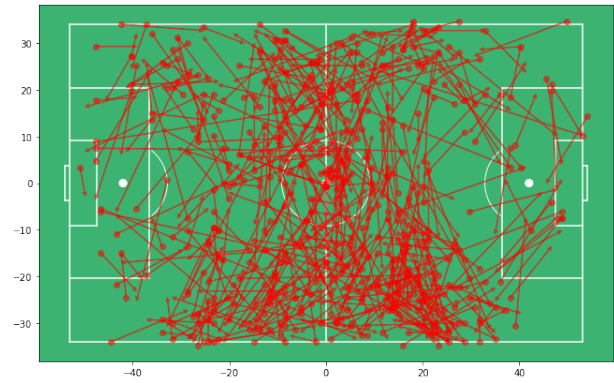


Figura 4.30: Pases completados del equipo local

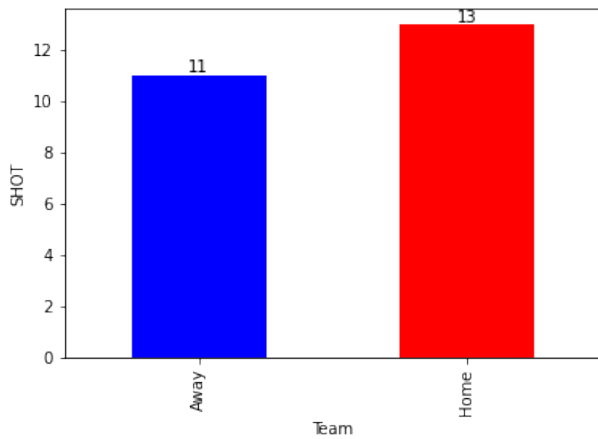


Figura 4.31: Tiros totales

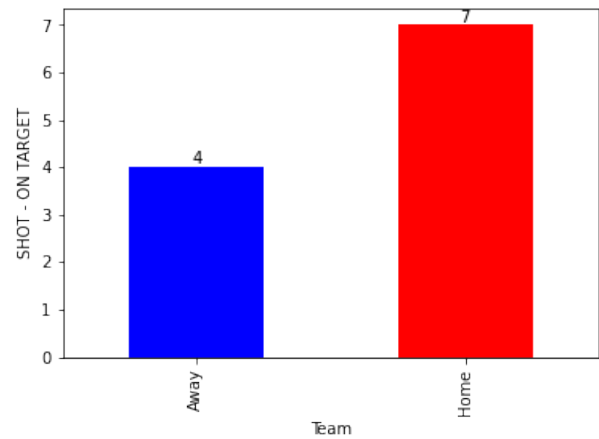


Figura 4.32: Tiros a puerta

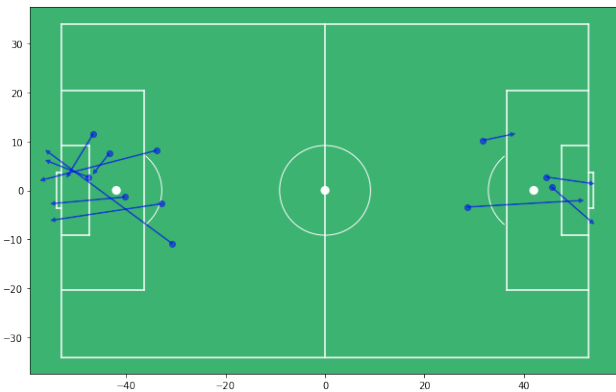


Figura 4.33: Tiros del equipo visitante

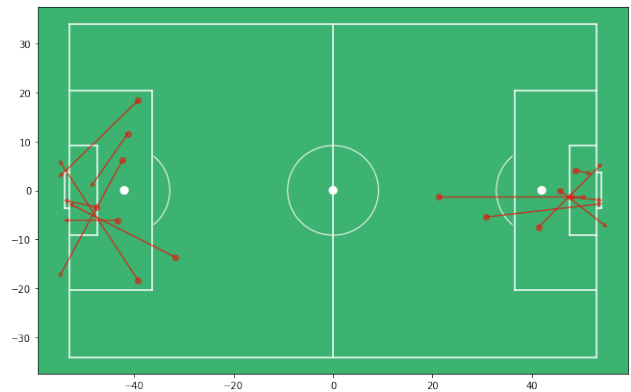


Figura 4.34: Tiros del equipo local

### 4.3.5. Visualizaciones avanzadas

Matplotlib nos ofrece una librería con la que crear animaciones llamada *animation*, con la cual podemos crear vídeos de como se van moviendo los jugadores por el campo. Hemos creado una función clip que recibe como parámetros los datos de las posiciones del equipo local y visitante y los fotogramas a reproducir. A modo de ejemplo ponemos a disposición un enlace a una carpeta de Google Drive para ver los goles del encuentro. [https://drive.google.com/drive/folders/18\\_Fc3zH\\_Ysgy46dqkZbmNFsprLlDuICL?usp=sharing](https://drive.google.com/drive/folders/18_Fc3zH_Ysgy46dqkZbmNFsprLlDuICL?usp=sharing)

## Pitch Control

Pitch control es un modelo desarrollado por William Spearman [6] [7], encargado de la ciencia de datos del Liverpool F.C (Subcampeón de la Premier League) que nos indica la probabilidad de que un equipo controle el balón en algún lugar del terreno de juego. Es decir, si tuviéramos el frame (Figura 4.24), podríamos saber la probabilidad que tendrían ambos equipos para controlar la posesión de balón. Esta probabilidad viene reflejada en forma de malla por encima del campo, mostrando en cada píxel sobre el terreno de juego un color, azul, si hay mas probabilidades que lo controle el equipo azul, o rojo, si el equipo rojo tiene más posibilidades. Para calcular esta probabilidad por todo el terreno tenemos asumir una serie de cuestiones como:

- Velocidad máxima de un jugador:  $5 \text{ m/s}$ .
- Aceleración máxima de un jugador:  $7 \text{ m/s}^2$ .
- Los jugadores eligen el camino mas corto hacia el balón (linea recta).
- Los jugadores tienen un tiempo de reacción de  $0,7 \text{ s}$ . Luego van hacia el balón a  $5 \text{ m/s}$ .

En el modelo de Spearman se tiene en cuenta el propio control del balón, el cual se tiene en cuenta como un proceso estocástico con una tasa fija. Para cada intervalo de tiempo  $\Delta t$  que un jugador esta cerca del balón, este tiene una probabilidad de  $\lambda * \Delta t$  de controlarlo. Según el modelo de Spearman  $\lambda = 4,30$ . Además, Spearman utiliza un función sigmoide para calcular la probabilidad de interceptar el balón, donde  $\sigma = 0,45$

Para realizar el calculo primero se crea un malla de píxeles proporcional a las medidas del terreno de juego, luego, se itera sobre cada una de los píxeles de la malla para poder calcular la probabilidad en cada píxel o punto. Para cada punto se itera sobre cada jugador en el terreno de juego, obteniendo los más próximos y calculando cual es la probabilidad de que llegue un jugador del equipo rojo o azul.

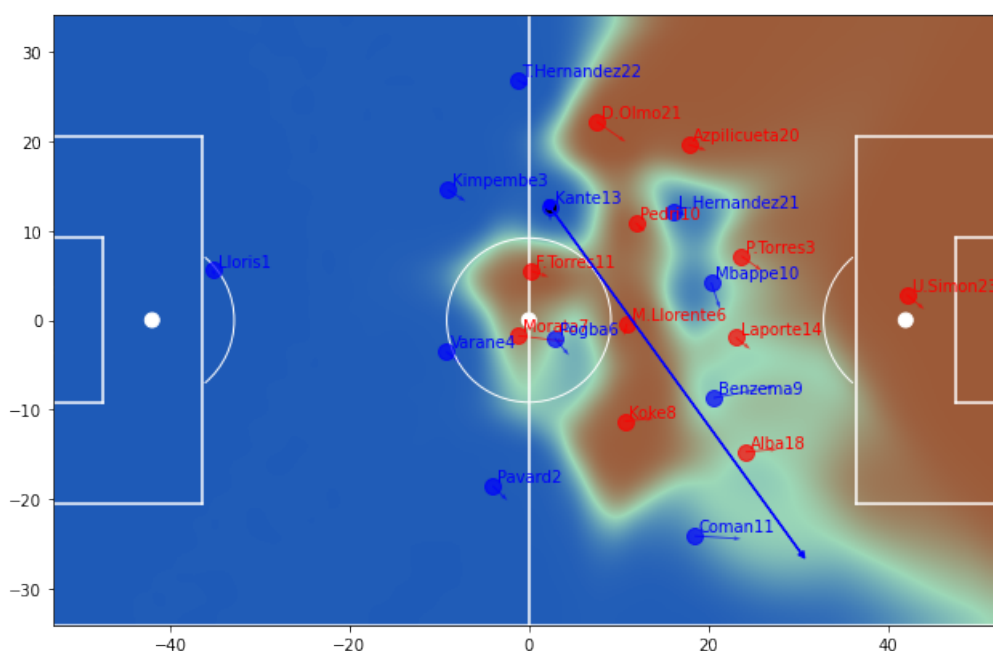


Figura 4.35: Ejemplo de Pitch Control

En la figura 4.35 podemos ver como el jugador número 13 efectúa un pase al jugador 11, en el mapa vemos que las posibilidades de que este controle el balón son bastante altas. A continuación podemos ver mas ejemplos, incluiremos las asistencias de gol del encuentro:

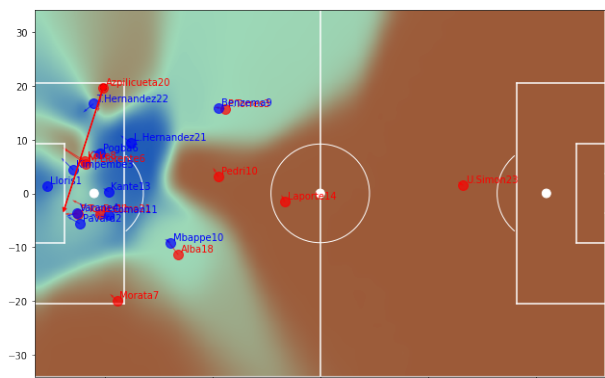


Figura 4.36: Asistencia del primer gol

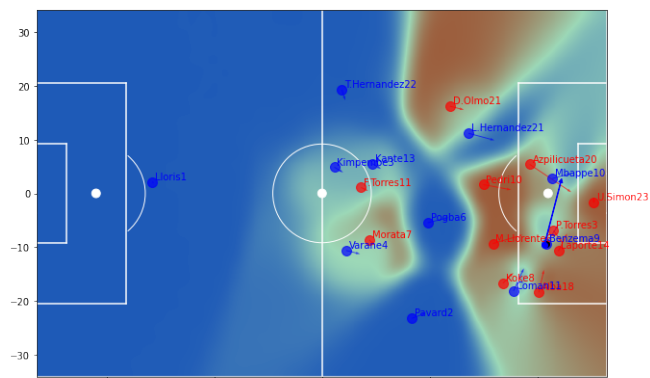


Figura 4.37: Asistencia del segundo gol

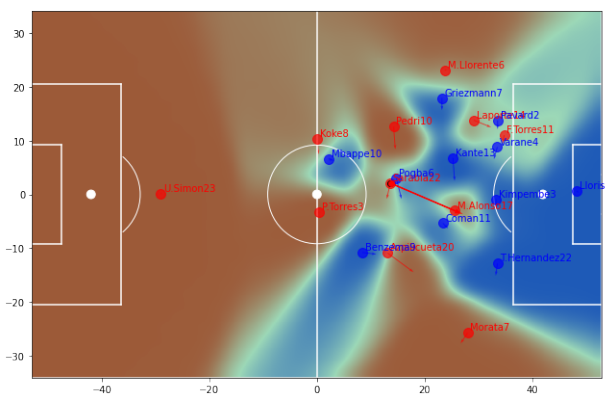


Figura 4.38: Asistencia del tercer gol

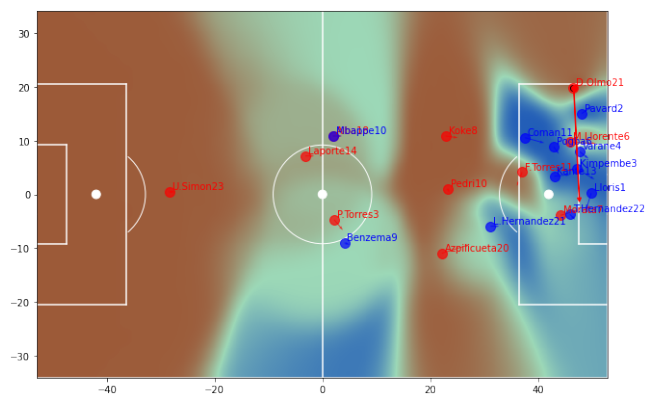


Figura 4.39: Asistencia del cuarto gol

De los 4 goles 3 fueron del equipo local (España), en los mapas de dichos goles podemos apreciar como la defensa del equipo visitante está bastante replegada, con casi todos sus hombres en el área. Es por esto que las probabilidades de controlar el balón para el equipo rojo eran reducidas, pero igualmente consiguieron anotar, esto demuestra muy poca solidez defensiva por parte del equipo azul.

Además, podemos hacer una clasificación de probabilidad de todos los pases que realizó cada equipo:

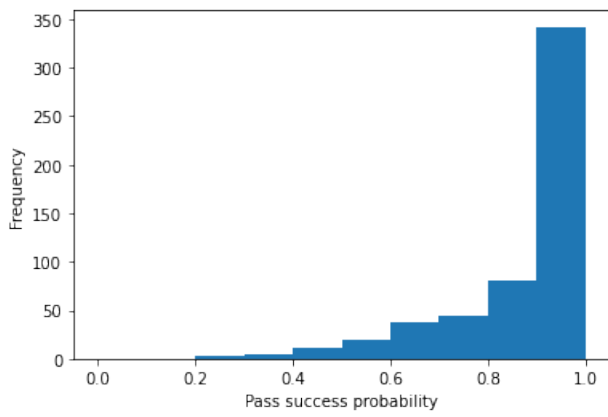


Figura 4.40: Clasificación de los pases del equipo local

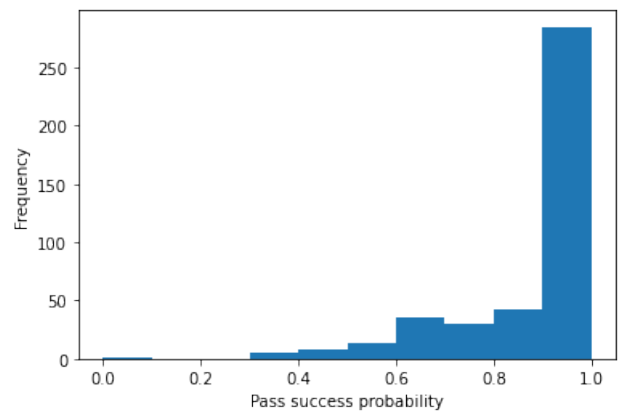


Figura 4.41: Clasificación de los pases del equipo visitante

# 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 diferentes tipos de datos para agilizar y ayudar al entendimiento de aspectos relevantes del fútbol, como son las lesiones y la anatomía de los jugadores, así como las disposición táctica de los equipos, movimientos, ocasiones, estadísticas, etc. de un partido.

El uso de elementos gráficos es algo vital para poder obtener un gran detalle sobre lo que puede suceder en el campo, es probable que ciertas gráficas, analizadas individualmente no aporten demasiado conocimiento, pero cuando hacemos un análisis en conjunto podemos sacar importantes conclusiones sobre el rendimiento del equipo o de algún jugador individual.

Por otro lado, la categoría de juego dentro de cualquier tipo de competición, es indiferente en cuanto a riesgo de lesiones. Sin embargo, la preparación física durante la temporada es relevante ya que puede influir en el índice lesional, por lo que es importante una buena programación y preparación.

Es recomendable que los equipos obtengan información detallada sobre la anatomía de sus jugadores, esto ayuda significativamente a la prevención de lesiones de los mismos. Vemos que existen una gran correlación entre las lesiones y los factores de riesgo que tiene cada uno de los integrantes de la plantilla, por eso es muy conveniente que los equipos inviertan más en una buena preparación física así como en programas para corregir o reducir estos factores de riesgo.

Cualquier equipo o entidad deportiva, independientemente de su nivel o recursos debería poder tener acceso a este tipo de herramientas y aprovecharlas para mejorar el rendimiento de sus jugadores

### 5.2. Líneas futuras

En este trabajo se han establecido las bases para continuar con el desarrollo de técnicas analíticas dentro del fútbol. El trabajo a continuar se centra en la incorporación de nuevas técnicas de análisis de datos más avanzadas como la IA o el ML, para poder obtener información aun mas detallada y profunda, así como obtener análisis predictivos, etc.

# Capítulo 6

## Summary and Conclusions

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

A working methodology was obtained with different kind of data to help understand relevant aspects of football, such as injuries and the anatomy of the players, as well as the tactical disposition of the teams, movements, occasions, statistics, etc. . of a football game.

The use of graphic elements is something vital to be able to obtain great detail about what can happen in the field, it is probable that certain graphs, analyzed individually, do not provide much knowledge, but when we do an analysis as a whole we can draw important conclusions about performance about the team or an individual player.

On the other hand, the category within any type of competition is indifferent in regarding the risk of injury. However, the physical preparation during the season is relevant since it can influence the injury rate, so good planning and preparation is important.

It is recommended that teams obtain detailed information about the anatomy of their players, this significantly helps prevent injuries. We see that there is a great correlation between injuries and the risk factors that each of the members of the squad has, that is why it is very convenient for the teams to invest more in good physical preparation as well as in programs to fix or reduce these factors.

Any sports team or entity, regardless of their level or resources, should be able to have access to this type of tools and take advantage of them to improve the performance of their players.

# Capítulo 7

## Presupuesto

En este capítulo presentamos una estimación del coste de la realización de este trabajo.

### 7.1. Presupuesto para el personal

En este apartado podemos ver el coste total de cada tarea que se ha realizado en este trabajo. El precio de cada tarea está calculado según las horas que los científicos de datos dediquen a cada tarea. Estos costes a priori son fijos pero pueden variar según el tiempo que les ocupen programar los algoritmos, las visualizaciones y demás configuraciones.

Tarea	Coste
Planteamiento de la metodología	200€
Obtención de los datos	300€
Procesamiento de los datos	500€
Creación de los algoritmos	300€
Visualización de los datos	250€
Total	1250€

Tabla 7.1: Presupuesto de las tareas

### 7.2. Presupuesto material

Por otro lado hay que tener en cuenta el aspecto material, ya que para estos estudios se necesitan una serie de herramientas como cámaras capaces de detectar movimiento, un ordenador para manejar los datos, medidores, etc.

Producto	Coste
Ordenador	900€
Programas y Librerías	0€
Materiales para medición de factores de riesgo	100€
Cámaras para registrar posiciones y eventos	2000€
Total	3000€

Tabla 7.2: Presupuesto material

### 7.3. Presupuesto Total

Una vez calculado el precio de cada apartado obtenemos el presupuesto final:

<b>Producto</b>	<b>Coste</b>
Presupuesto para personal	1250€
Presupuesto material	3000€
Presupuesto total	3250€

Tabla 7.3: Presupuesto material



# Apéndice A

## Algoritmos del caso de lesiones

En este apéndice se incluyen diversos algoritmos utilizados para la modificación de los datos y visualizaciones del caso de las lesiones. Todo el código del trabajo se puede encontrar en este repositorio: <https://github.com/EindhovenLion99/TFG-DataScience>

### A.1. Algoritmo para separar el vector de lesiones en diferentes celdas

```
/*
 *
 * Fichero DataFunctions.py
 *
 *****/
def label_race(row):
    print(row['Lesiones Previas'])
    return row['Lesiones Previas'] + " ; " + row['Lesiones Actuales']

def getInjuriesTable(jugadores, periodo_lesion, COMBO=False):
    lesiones_previas = jugadores[['Equipo', 'Edad', 'Altura', periodo_lesion]]
    if COMBO:
        lesiones_previas = jugadores[['Equipo', 'Edad', 'Altura', periodo_lesion, 'Lesiones Actuales']]
        lesiones_previas[periodo_lesion] = lesiones_previas.apply(lambda row: label_race(row), axis=1)
        lesiones_previas['Vector Lesiones'] = lesiones_previas[periodo_lesion].str.split(" ; ")
        lesiones_previas = lesiones_previas.explode('Vector Lesiones')
        lesiones_previas['Lesion'] = lesiones_previas['Vector Lesiones'].str.split("-").str[0]
        lesiones_previas['Parte'] = lesiones_previas['Vector Lesiones'].str.split("-").str[1]
        lesiones_previas['Grupo Muscular'] = lesiones_previas['Vector Lesiones'].str.split("-").str[-1]
        lesiones_previas = lesiones_previas.drop(columns=[periodo_lesion])
    return lesiones_previas

*****/
```

### A.2. Algoritmo para visualizar las lesiones de cada equipo

```

/*****
def plotNumLesionesPorEquipo(jugadores, tipo):
    team_total_injuries = jugadores.groupby(['Equipo']).sum()[tipo].sort_values()
    print(team_total_injuries)
    pl_mean = team_total_injuries.mean()
    ax_2 = team_total_injuries.plot(kind = "barh")
    ax_2.set_xticks(np.arange(0, max(team_total_injuries) + 2, 2))
    ax_2.set(xlabel="Lesiones", title="Numero de lesiones por equipo")
    ax_2.axvline(pl_mean, ls="--", color='r')
    ax_2.set_title("Lesiones por equipo (" + tipo + ")")
*****/

```

### A.3. Algoritmo para mostrar las lesiones combinando factores de riesgo

```

/*****
*
* Fichero plotFunctions.py
*
*****/
def label_race(row, factores):
    count = 0
    for fact in factores:
        if row[fact] == 'Si':
            count += 1
    if count == len(factores):
        return 'Si'
    else:
        return 'No'

def plotFactorRiesgoCombinado(jugadores, factores, zona, figsize=(10,15)):
    new_string = ""
    for factor in factores:
        new_string += factor + " "
        jugadores[factor] = ['Si' if valor != 'No'
                               and valor != 'Normal'
                               and valor != 'Neutro'
                               else valor for valor in jugadores[factor]]
    jugadores[new_string] = jugadores.apply(lambda row: label_race(row, factores), axis=1)
    pl = jugadores.groupby([zona, new_string])[new_string].count().unstack()
    ax_ = pl.plot(figsize=figsize, kind = "barh", stacked = True)
    for c in ax_.containers:
        labels = [int(v.get_width()) if v.get_width() > 0 else '' for v in c]
        ax_.bar_label(c, labels=labels, label_type='center')
*****/

```

# Apéndice B

## Algoritmos del caso de data-tracking

En este apéndice se incluyen diversos algoritmos utilizados para la modificación de los datos y visualizaciones del caso de data-tracking.

### B.1. Algoritmo para modificar los datos de las posiciones

```
/******  
*  
* Fichero DataFunctions.py  
*  
*****  
def getGamePositions(game_id, team, DATADIR):  
    eventfile = DATADIR + '/Sample_Game_%d/Sample_Game_%d_RawTrackingData_%s_Team.csv'  
                % (game_id, game_id, team)  
    csvfile = open(eventfile, 'r')  
    reader = csv.reader(csvfile)  
    team = next(reader)[3].lower()  
  
    team_numbers = [number for number in next(reader) if number != '']  
    columns = next(reader)  
  
    for i, j in enumerate(team_numbers):  
        columns[i * 2 + 3] = team + "_" + j + "_x"  
        columns[i * 2 + 4] = team + "_" + j + "_y"  
  
    columns[-2] = "ball_x"  
    columns[-1] = "ball_y"  
  
    positions = pd.read_csv(eventfile, names = columns, index_col = 'Frame', skiprows = 3)  
    positions = to_metric_coordinates(positions)  
    positions = getPlayerVel(positions)  
    return positions  
  
def getPlayerVel(team, maxspeed = 12):  
    player_ids = np.unique( [ c[:-2] for c in team.columns if c[:4] in ['home','away'] ] )  
    dt = team['Time [s]'].diff()  
    for player in player_ids: # cycle through players individually  
        # difference player positions in timestep dt to get unsmoothed estimate of velocity  
        vx = team[player + "_x"].diff() / dt
```

```

vy = team[player + "_y"].diff() / dt

if maxspeed > 0:
    raw_speed = np.sqrt( vx**2 + vy**2 )
    vx[ raw_speed > maxspeed ] = np.nan
    vy[ raw_speed > maxspeed ] = np.nan

    team[player + "_vx"] = vx
    team[player + "_vy"] = vy
    team[player + "_speed"] = np.sqrt( vx**2 + vy**2 )
return team
*****/

```

## B.2. Algoritmo para calcular la velocidad de los jugadores

```

/*****
def getPlayerVel(team, maxspeed = 12):
    player_ids = np.unique( [ c[:-2] for c in team.columns if c[:4] in ['home','away'] ] )
    dt = team['Time [s]'].diff()
    for player in player_ids: # cycle through players individually
        # difference player positions in timestep dt to get unsmoothed estimate of velocity
        vx = team[player + "_x"].diff() / dt
        vy = team[player + "_y"].diff() / dt

        if maxspeed > 0:
            raw_speed = np.sqrt( vx**2 + vy**2 )
            vx[ raw_speed > maxspeed ] = np.nan
            vy[ raw_speed > maxspeed ] = np.nan

            team[player + "_vx"] = vx
            team[player + "_vy"] = vy
            team[player + "_speed"] = np.sqrt( vx**2 + vy**2 )
    return team
*****/

```

## B.3. Algoritmo para mostrar los jugadores en el campo

```

/*****
*
* Fichero plotFunctions.py
*
*****/
def plot_frame(home_team, away_team, colors=('r', 'b'), PlayerMarkerSize=10,
              PlayerAlpha=0.7, annotate=False, velocity=False, figax=None):
    if figax is None: # create new pitch
        fig,ax = plot_field()
    else: # overlay on a previously generated pitch

```

```

fig,ax = figax # unpack tuple
for team,color in zip( [home_team, away_team], colors):
    x_columns = [c for c in team.keys() if c[-2:].lower() == '_x' and c != 'ball_x']
    y_columns = [c for c in team.keys() if c[-2:].lower() == '_y' and c != 'ball_y']
    ax.plot(team[x_columns], team[y_columns], color + 'o', markersize=PlayerMarkerSize,
            alpha=PlayerAlpha)
    if velocity:
        vx_columns = ['{}_vx'.format(c[:-2]) for c in x_columns]
        vy_columns = ['{}_vy'.format(c[:-2]) for c in y_columns]
        ax.quiver(team[x_columns], team[y_columns], team[vx_columns], team[vy_columns],
                 color=color, scale_units='inches', scale=10, width=0.0015,
                 headlength=5, headwidth=3, alpha=PlayerAlpha)
    if annotate:
        [ ax.text(team[x]+0.5, team[y]+0.5, x.split('_')[1], fontsize=10, color=color )
          for x,y in zip(x_columns,y_columns) if not (np.isnan(team[x]) or np.isnan(team[y]))]
    ax.plot(home_team['ball_x'], home_team['ball_y'], 'ko', alpha=1.0)
return fig, ax
*****/

```

# Siglas

**BD** Big Data 3, 4

**IA** Inteligencia Artificial 1, 3, 4, 26

**ML** Machine Learning 3, 4, 26

# Bibliografía

- [1] Inderpal Bhandari, Edward Colet, Jennifer Parker, Zachary Pines, Rajiv Pratap, and Krishnakumar Ramanujam. Advanced scout: Data mining and knowledge discovery in nba data. *Data Mining and Knowledge Discovery*, 1(1):121–125, 1997.
- [2] Christoph Buck and Sebastian Ifland. Toward an enduring football economy: a business model taxonomy for europe’s professional football clubs. *European Sport Management Quarterly*, 0(0):1–21, 2022.
- [3] John. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(03):90–95, may 2007.
- [4] Wes McKinney. pandas: a foundational python library for data analysis and statistics. *Python for high performance and scientific computing*, 14(9):1–9, 2011.
- [5] Keshav Rathi, Priyam Somani, Aditya V Koul, and KS Manu. Applications of artificial intelligence in the game of football: The global perspective. *Researchers World*, 11(2):18–29, 2020.
- [6] William Spearman. Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference*, pages 1–17, 2018.
- [7] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.
- [8] Poojan Thakkar and Manan Shah. An assessment of football through the lens of data science. *Annals of Data Science*, 8(4):823–836, 2021.
- [9] William Vallejo, Carlos Díaz-Uribe, and Catalina Fajardo. Google colab and virtual simulations: Practical e-learning tools to support the teaching of thermodynamics and to introduce coding to students. *ACS omega*, 7(8):7421–7429, 2022.
- [10] Nicholas M Watanabe, Stephen Shapiro, and Joris Drayer. Big data and analytics in sport management. *Journal of Sport Management*, 35(3):197–202, 2021.