



Universidad de Valladolid



**ESCUELA DE INGENIERÍAS
INDUSTRIALES**

UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERIAS INDUSTRIALES

Grado en Ingeniería en Organización Industrial

**IMPLEMENTACIÓN DE UNA HERRAMIENTA DE
ANÁLISIS EN LA GESTIÓN DEPORTIVA
MEDIANTE POWER BI**

Autor:

De Miguel Toquero, Alberto

Tutor:

Villafáñez Cardeñoso, Félix Antonio

Departamento de Organización de Empresas y
Comercialización e Investigación de Mercados

Valladolid, septiembre de 2020.





PRÓLOGO

Hasta llegar a primero de bachillerato, todos los estudiantes cursamos las mismas asignaturas y damos prácticamente el mismo temario. Sin embargo, cada uno las percibe de forma diferente. Algunos sienten pasión por las matemáticas, la física o el dibujo técnico y otros son amantes de la lengua y la literatura, los idiomas, la historia y la filosofía.

Normalmente, ese suele ser el factor que marque la tan importante decisión de elegir el bachillerato de ciencias o el de letras. Otras veces son los padres los que fuerzan a coger uno u otro. En mi caso, las matemáticas y la física eran dos de mis asignaturas favoritas. Cuando a medida que avanzaba el curso iba viendo como progresaba en matemáticas y resolvía con facilidad esos problemas que un par de meses atrás resultaban imposibles, la sensación era bastante gratificante. En física la sensación era diferente. Eran el asombro, la fascinación y la magnitud de las cosas que se podían lograr dominando la física las que lograban que me gustara esa materia.

Sin embargo, al mismo tiempo, me encantaban la literatura y, especialmente, la filosofía y la historia (en concreto la Historia de España). Por ello, para decidirme por uno u otro bachillerato tuve que emplear un razonamiento diferente al de la mayoría de la gente. Cogí la opción de ciencias porque me dijeron que era más fácil cursar una carrera de letras tras hacer el bachillerato de ciencias que lo contrario.

Al acabar el bachillerato dudaba entre estudiar derecho e ingeniería. De nuevo, me basé en el mismo principio para decidir. Si entraba a derecho y me arrepentía, sería difícil coger el ritmo de las asignaturas de ingeniería después de un año sin haber dado ninguna materia de esa rama. En el caso contrario no había ese problema.

Queda bastante claro que no fue la pasión por la tecnología el principal argumento que me llevara a estudiar esta carrera universitaria. Pese a ello, jamás he considerado que fuera un error entrar en el mundo de la ingeniería.

Este campo es infinitamente grande y complejo. En primero dimos las materias troncales o básicas, comunes a todas las titulaciones de ingeniería. Matemáticas, física, expresión gráfica, química, etc.

En segundo ya entramos de pleno en el ámbito ingenieril. Fundamentos de electrónica, ingeniería de organización, resistencia y ciencia de los materiales, ingeniería fluidomecánica, fundamentos de automática y un largo etcétera de materias que tocaban las diversas ramas existentes.

Entre otras cosas, esta organización de asignaturas está hecha de este modo para que un alumno pueda ver qué es lo que más le gusta y pueda cambiarse de



grado, dentro de la Escuela, para en tercero y cuarto cursar aquella especialidad o rama que más le atraiga.

En mi caso escogí la Ingeniería en Organización Industrial, porque me parecía que mezclaba de un modo interesante y útil el mundo de la ingeniería, la ciencia y la tecnología, por un lado, y el mundo de la empresa, la economía y la gestión por otro.

En tercero y cuarto he dado asignaturas como dirección de empresas, gestión de la innovación y creación de empresas, dirección estratégica, dirección de proyectos, TICS, estadística empresarial, métodos cuantitativos en ingeniería de organización... donde efectivamente hemos visto cómo aplicar los conceptos técnicos al mundo empresarial real. Y aquí llega la razón de ser de este trabajo.

Está claro que para construir un edificio se necesitan unos conocimientos muy potentes de ciencia e ingeniería de los materiales, de resistencia y elasticidad de materiales y de electricidad o climatización entre otras áreas. Para diseñar y fabricar un coche hay que saber mucha electrónica y para diseñar el sistema de tuberías de una planta industrial hay que conocer cómo se comportan los fluidos, además de emplear complejos programas de diseño asistido por ordenador. Estos son los ingredientes principales, pero para que el producto final triunfe es imprescindible un condimento: la información.

La información es el arma más poderosa que tiene la humanidad en la actualidad. Hoy en día, las empresas y los gobiernos de todo el mundo se gastan millones y millones de euros en obtener datos. Datos sobre personas, sobre otras empresas y otros gobiernos, sobre una línea de producción... sobre absolutamente todo.

De nada servirá conocer todas las propiedades de un material si no tenemos la información (los datos) de cómo ha funcionado previamente al hacer aquello para lo que lo queramos usar. Teniendo el conocimiento de base, teniendo los datos y sabiendo extraer la información de ellos, no existe prácticamente nada que no se pueda lograr.

Este TFG pretende mostrar cómo se puede coger un conjunto muy grande de datos, mareantes para cualquier persona, y aplicándoles el análisis adecuado se pueden extraer conclusiones valiosas, rápidas y sencillas de interpretar.

El caso práctico sobre el que se desarrollará es la gestión deportiva, lo cual no tiene especial relación con la ingeniería. Sin embargo, el valor y el interés del análisis de datos está en que el mismo procedimiento y el mismo método podrían servir con datos sobre una línea de producción, datos sobre una red logística, datos sobre el comportamiento celular y molecular de pacientes médicos con una determinada patología o enfermedad; y así podría seguir rellenando hojas y hojas con ejemplos reales y útiles para la sociedad en los que el análisis de datos es un ingrediente imprescindible.



RESUMEN

Este trabajo presenta cómo, mediante técnicas de Análisis de Datos y la aplicación de la Inteligencia de Negocio, podemos llegar a conocer de forma exhaustiva diferentes aspectos del funcionamiento de una organización.

Se comienzan desarrollando los conceptos de Análisis de Datos y Big Data, analizando sus definiciones, características, evolución histórica y su posible futuro. Se detallan también algunos conceptos relacionados, como la segmentación de datos, bases de datos, inteligencia artificial o KPIs, entre otros, hasta llegar a los conceptos de Inteligencia de Negocio, la Minería de Datos y los Tableros de Visualización.

Finalmente se presenta un caso práctico: “un equipo de fútbol desea conocer el rendimiento de sus jugadores para, a partir de los resultados, decidir a quien se debe renovar o vender y a quien fichar, comparando a sus futbolistas con los de otros equipos”, y se describe cómo se ha desarrollado, empleando Power BI, una respuesta a esta necesidad.

Palabras clave

Análisis de datos, Big Data, Inteligencia de Negocio, Power BI, Fútbol

ABSTRACT

This project aims to demonstrate how, through data analysis and the application of business intelligence, we can know exhaustively how an organization works. A practical case is presented: a soccer team wants to know the performance of its players and, from the results, decide who should be renewed and sold and who should be signed, comparing their footballers with those of other teams.

Previously, the concepts Data Analysis and Big Data have been developed, studying some definitions and their characteristics; and exposing its historical evolution and its possible future development. Some concepts such as data segmentation, databases, artificial intelligence or KPIs, among others, have been analyzed.

Business Intelligence has been studied, where data mining and dashboards have been discussed. The program used in the practical assumption, Power BI, is also presented.

Keywords

Data analysis, Big Data, Business Intelligence, Power BI, Football, Soccer





Contenido

PRÓLOGO	3
RESUMEN.....	5
Palabras clave.....	5
ABSTRACT.....	5
Keywords	5
Contenido.....	7
ÍNDICE DE SIGLAS Y ACRÓNIMOS	11
ÍNDICE DE FIGURAS.....	13
ÍNDICE DE TABLAS.....	15
1. INTRODUCCIÓN Y OBJETIVOS	17
1.1. CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO	17
1.2. OBJETIVOS.....	17
1.3. ESTRUCTURA DEL DOCUMENTO.....	18
2. ANÁLISIS DE DATOS Y BIG DATA	21
2.1. DEFINICIONES Y CARACTERÍSTICAS	21
2.1.1. ¿Qué es el Big Data y cuáles son sus características?.....	21
2.1.2. ¿Qué es el análisis de datos y cuáles son sus características, objetivos y ventajas?.....	26
2.2. HISTORIA DEL ANÁLISIS DE DATOS Y EL BIG DATA	29
2.2.1. Origen, evolución y situación actual.....	29
2.2.2. El futuro. Posible evolución y debates abiertos	32
2.2.3. Historia del Análisis de Datos y el Big Data en el fútbol	34
2.3. CAMPOS DE APLICACIÓN: ALGUNOS EJEMPLOS	37
2.3.1. Medicina.....	37
2.3.2. Agricultura	38
2.3.3. Comunicación y periodismo.....	40
2.3.4. Desarrollo local y sector público. Países en vías de desarrollo.....	41
2.3.5. Turismo.....	42
2.4. BIG DATA Y ANÁLISIS DE DATOS EN EL DEPORTE.....	43
2.4.1. El futuro del Big Data en el fútbol	45
2.5. CONCEPTOS ASOCIADOS AL BIG DATA Y AL ANÁLISIS DE DATOS	46
2.5.1. Small Data.....	47



2.5.2.	Datos Abiertos (<i>Open Data</i>)	48
2.5.3.	Datos sin Procesar (Raw Data).....	49
2.5.4.	Tipos de bases de datos: Relacional y no relacional	49
2.5.5.	Tipos de recogida de datos: Clasificación IBM.....	52
2.5.6.	Árboles de decisión	54
2.5.7.	ETL.....	56
2.5.8.	Inteligencia Artificial	57
2.5.9.	Industria 4.0.....	59
2.5.10.	Efecto Moneyball	60
2.5.11.	Almacenes de datos (Data Warehouses).....	61
2.5.12.	Segmentación de Datos.....	62
2.5.13.	Raspado Web.....	63
2.6.	KPIs.....	64
3.	INTELIGENCIA DE NEGOCIO.....	69
3.1.	¿QUÉ ES LA INTELIGENCIA DE NEGOCIO?.....	69
3.2.	OBJETIVOS, CARACTERÍSTICAS Y VENTAJAS	70
3.3.	MINERÍA DE DATOS.....	72
3.4.	TABLEROS DE VISUALIZACIÓN	73
4.	POWER BI.....	77
4.1.	OTRAS HERRAMIENTAS EXISTENTES	78
4.1.1.	Pentaho	78
4.1.2.	Tableau.....	79
4.1.3.	IBM Watson Analytics.....	79
4.1.4.	Apache Spark.....	80
5.	CASO PRÁCTICO.....	81
5.1.	PRESENTACIÓN DEL SUPUESTO	81
5.1.1.	¿Quién es el cliente?	81
5.1.2.	¿Qué quiere el cliente?	83
5.1.3.	¿Cómo se va a desarrollar la herramienta de análisis en la gestión deportiva?.....	85
5.2.	IMPLEMENTACIÓN DE LA HERRAMIENTA EN POWER BI	88
5.2.1.	Carga de las tablas en Power BI.....	88
5.2.2.	Desarrollo y tratamiento de la tabla de estadísticas de los jugadores.....	95
5.2.3.	Creación del informe sobre el análisis de los futbolistas.....	104



6. RESULTADOS OBTENIDOS Y CONCLUSIONES.....	129
6.1. VALORACIÓN DE LOS RESULTADOS OBTENIDOS Y DEL GRADO DE CONSECUCIÓN DE LOS OBJETIVOS MARCADOS.....	129
6.2. CONCLUSIONES EXTRAÍDAS DE LA ELABORACIÓN DE ESTE TRABAJO....	130
7. BIBLIOGRAFÍA.....	133





ÍNDICE DE SIGLAS Y ACRÓNIMOS

BBDD: Bases de datos.

BI: Inteligencia de Negocio o Inteligencia Empresarial (BI, por sus siglas en inglés, *Business Intelligence*).

CD: Club Deportivo.

CF: Club de Fútbol.

CSV: Valores separados por comas (CSV, por sus siglas en inglés, *comma-separated values*).

EPA: Encuesta de población activa.

ERP: Sistemas de planificación de recursos empresariales (ERP, por sus siglas en inglés, *Enterprise Resource Planning*).

ETL: Extraer, transformar y cargar (ETL, por sus siglas en inglés, *Extract, Transform and Load*).

FIFA: Federación Internacional de Fútbol Asociación (FIFA, por sus siglas en francés, *Fédération Internationale de Football Association*).

HCDSNS: Historia Clínica Digital del Sistema Nacional de Salud.

IA: Inteligencia Artificial.

IoT: Internet de las Cosas (IoT, por sus siglas en inglés, *Internet of Things*).

IPC: Índice de precios de consumo o Índice de precios al consumidor.

KPI: Indicador clave de rendimiento (KPI, por sus siglas en inglés, *key performance indicator*).

LFP: Liga de Fútbol Profesional [de España].

MRP: Planificación de los requerimientos de material (MRP, por sus siglas en inglés, *Material Requirements Planning*).

NBA: Asociación Nacional de Baloncesto [de Estados Unidos] (NBA, por sus siglas en inglés, *National Basketball Association*).

NFL: Liga Nacional de fútbol americano [de Estados Unidos] (NFL, por sus siglas en inglés, *National Football League*).

ONG: Organización no Gubernamental.

ROI: Retorno de la inversión (ROI, por sus siglas en inglés, *Return On Investment*).



S.A.D.: Sociedad Anónima Deportiva.

SD: Sociedad Deportiva.

SNS: Sistema Nacional de Salud de España.

SQL: Lenguaje de consulta estructurada (SQL, por sus siglas en inglés, *Structured Query Language*).

TFG: Trabajo de fin de grado.

TIC: Tecnologías de la información y la comunicación.

UD: Unión Deportiva.

UEFA: Unión de Federaciones Europeas de Fútbol (UEFA, por sus siglas en francés, *Union des Associations Européennes de Football*).

XML: Lenguaje de Mercado Extensible (XML, por sus siglas en inglés, *eXtensible Markup Language*).



ÍNDICE DE FIGURAS

Ilustración 1. Datos generados en Internet en un minuto. Fuente: (Trabado, 2018)	22
Ilustración 2. Las 7 uves del Big Data. Fuente: (Big Data International Campus, 2020).	25
Ilustración 3. Agricultura de Precisión con Imágenes de Satélite, Sensorización y Big Data. Fuente: (Feval Institución Ferial de Extremadura, 2018).	39
Ilustración 4. Modelo de flujograma algorítmico aplicado a los medios. Fuente: (Flores Vivar, 2018).	40
Ilustración 5. Esquema de una Base de Datos Relacional. Fuente: (TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN II (TIC'S II), s. f.).	52
Ilustración 6. Ejemplo de árbol de decisión. Fuente: (Martínez & Ventura, 2016) ..	56
Ilustración 7. Fases del análisis de datos de la Inteligencia de Negocio. Fuente: (Matesanz Niño, 2018) ..	69
Ilustración 8. Pantalla del Editor Avanzado de Power Query. Fuente: Elaboración Propia ..	90
Ilustración 9. Pantalla del Editor Avanzado de Power Query una vez se han llevado a cabo todas las modificaciones indicadas. Fuente: Elaboración Propia ..	91
Ilustración 10. Pantalla donde definir la columna personalizada a agregar. Fuente: Elaboración Propia ..	92
Ilustración 11. Pantalla de la columna personalizada, con ésta ya definida. Fuente: Elaboración Propia ..	93
Ilustración 12. Pantalla del Editor de Power Query, una vez creada la columna donde aparece la tabla de cada equipo. Fuente: Elaboración Propia.....	94
Ilustración 13. Tabla donde aparecen las estadísticas de todos los jugadores de todos los equipos de primera, segunda y segunda B. Fuente: Elaboración Propia..	95
Ilustración 14. Creación de la columna "Minutos/GE". Fuente: Elaboración Propia	98
Ilustración 15. Creación de la columna "Minutos/TA". Fuente: Elaboración Propia.	99
Ilustración 16. Creación de la columna "Minutos/TR". Fuente: Elaboración Propia ..	100
Ilustración 17. Columna donde se calcula una valoración global del rendimiento de cada jugador. Fuente: Elaboración Propia.....	103
Ilustración 18. Portada del informe del análisis del rendimiento de los jugadores del Real Valladolid. Fuente: Elaboración Propia.....	105
Ilustración 19. Comparador de jugadores de campo. Fuente: Elaboración Propia	107
Ilustración 20. Resultado de editar las interacciones del filtro de la izquierda. Fuente: Elaboración Propia.....	109
Ilustración 21. Comparador de Porteros. Fuente: Elaboración Propia	111
Ilustración 22. Matrices Estratégicas I. Fuente: Elaboración Propia	112
Ilustración 23. Matrices Estratégicas II. Fuente: Elaboración Propia	114
Ilustración 24. Regla para establecer los colores de la columna "Minutos/G". Fuente: Elaboración Propia.....	115



Ilustración 25. Regla para establecer los colores de la columna "Min". Fuente: Elaboración Propia	115
Ilustración 26. Regla para establecer los colores de la columna "PS". Fuente: Elaboración Propia	115
Ilustración 27. Regla para establecer los colores de la columna "valoracion_jugadores". Fuente: Elaboración Propia	116
Ilustración 28. Tarjeta donde aparecen los centrocampistas del Real Valladolid, en orden decreciente de valoración obtenida. Fuente: Elaboración Propia.....	117
Ilustración 29. Vista de la página donde aparece la tabla semafórica y las tarjetas de jugadores mejor valorados por posición. Fuente: Elaboración Propia	118
Ilustración 30. Portada del informe sobre el análisis de los jugadores del resto de los equipos. Fuente: Elaboración Propia	118
Ilustración 31. Primera columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia	119
Ilustración 32. Segunda columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia.....	120
Ilustración 33. Tercera columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia	120
Ilustración 34. Primer paso llevado a cabo para asociar el código de un equipo con su nombre. Fuente: Elaboración Propia.....	121
Ilustración 35. Pantalla del Editor Avanzado de Power Query donde se ha programado la iteración para obtener el nombre de cada equipo. Fuente: Elaboración Propia	122
Ilustración 36. Columna agregada para mostrar el nombre del equipo. Fuente: Elaboración Propia	122
Ilustración 37. Comparador de jugadores del informe de análisis de todos los equipos. Fuente: Elaboración Propia	123
Ilustración 38. Comparador de porteros del informe de análisis de todos los equipos. Fuente: Elaboración Propia	123
Ilustración 39. Vista filtrada de la página "Matrices Estratégicas I" del segundo informe. Fuente: Elaboración Propia.....	124
Ilustración 40. Vista filtrada de la página "Matrices Estratégicas II" del segundo informe. Fuente: Elaboración Propia.....	125
Ilustración 41. Primera columna agregada para obtener el código de cada jugador. Fuente: Elaboración Propia.....	126
Ilustración 42. Segunda columna agregada para obtener el código de cada jugador. Fuente: Elaboración Propia.....	126
Ilustración 43. Tabla semafórica y comparativa de valoración por puestos del segundo informe. Fuente: Elaboración Propia	127
Ilustración 44. Tabla comparativa de equipos según su valoración por puestos. Fuente: Elaboración Propia.....	127



ÍNDICE DE TABLAS

Tabla 1. Tipos de datos. Fuente: (QuestionPro, 2018).....	27
Tabla 2. Cuando usar una Base de Datos relacional y no relacional. Fuente: (Alexis Rendón, 2019)	52
Tabla 3. Vista de la tabla de estadísticas de los jugadores del Real Valladolid en Power BI. Fuente: Elaboración Propia	96
Tabla 4. Leyenda con los términos de la tabla de estadísticas de los jugadores. Fuente: Elaboración Propia.....	97
Tabla 5. Leyenda de las posiciones de los jugadores. Fuente: Elaboración Propia .	97
Tabla 6. Tabla de estadísticas de los jugadores del Real Valladolid en la temporada 2019/2020, una vez se han agregado las columnas "Minutos/GE", "Minutos/TA" y "Minutos/TR". Fuente: Elaboración Propia	101
Tabla 7. Tabla semafórica de comparación del rendimiento de los jugadores del Real Valladolid. Fuente: Elaboración Propia.....	116





1. INTRODUCCIÓN Y OBJETIVOS

1.1. CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO

El Grado en Ingeniería en Organización Industrial, impartido en la Escuela de Ingenierías Industriales de la Universidad de Valladolid, tiene un plan de estudios de cuatro años de duración. Los dos primeros años son comunes al resto de grados de la Escuela, donde se ven asignaturas como física, matemáticas, química, estadística, expresión gráfica, ingeniería fluidomecánica, termodinámica técnica y transmisión del calor, electrotecnia, fundamentos de electrónica, ciencia de los materiales, ingeniería de organización y sistemas de producción y fabricación, entre otras.

Por el contrario, en el tercer y el cuarto año se imparten asignaturas específicas en cada grado, las correspondientes a su especialidad. En el caso concreto de la Ingeniería en Organización Industrial se estudian materias como dirección de proyectos, entorno económico y estrategia, dirección de empresas, métodos cuantitativos en ingeniería de organización, gestión de la innovación y creación de empresas, dirección estratégica, automatización industrial, diseño de sistemas productivos y logísticos, tics para la gestión de la empresa, etc.

Viendo el listado de materias enumeradas, tanto de los dos primeros años como de los dos últimos, queda patente que en este grado universitario se mezclan convenientemente dos ramas de estudio: la ciencia y la tecnología y, por otro lado, el mundo empresarial y la gestión industrial. En este punto, cabría hacerse la siguiente pregunta: ¿cuál es el nexo de unión entre ambas?

La respuesta es, a mi juicio, simple: los datos. La información.

El ingeniero en organización industrial es un perfil muy demandado en el mercado laboral actual. Sus conocimientos sobre aspectos técnicos son elevados. Sin embargo, no creo que esa sea su mayor valía. En mi opinión, ese honor le corresponde a su capacidad de, con los conocimientos que tiene, analizar los problemas que surgen, encontrar las posibles soluciones y elegir la mejor de ellas.

Todo ello sería imposible sin interpretar adecuadamente la información. Recabar los datos, estudiarlos y analizarlos para obtener información y dominar y comprender ésta para generar conocimiento, son los pasos necesarios para que las empresas adquieran ventajas competitivas sobre el resto.

Este Trabajo de Fin de Grado busca justificar todo lo que se acaba de exponer, empleando como marco la dirección o gestión deportiva de un equipo de fútbol.

1.2. OBJETIVOS

El objetivo principal de este TFG es implementar, mediante el programa Power BI, una herramienta de análisis en la gestión deportiva. Con ello, se pretende demostrar cómo, mediante un correcto análisis de datos y la aplicación de la Inteligencia de Negocio o Inteligencia Empresarial, se puede mejorar la gestión de la



plantilla de un equipo de fútbol y optimizar la toma de decisiones, algo que resultaría imposible de lograr si se pretendiera analizar los datos en bruto.

A mayores, se pueden citar una serie de objetivos complementarios, que servirán para profundizar más en las áreas de estudio que aborda este trabajo. Son los que se exponen a continuación:

- Conocer en profundidad qué son el análisis de datos y el Big Data y cuáles son sus características y diferencias.
- Conocer y analizar algunos de los debates que existen en torno a estos conceptos y proponer la posible evolución futura en el corto y medio plazo.
- Estudiar las ventajas y mejoras que el análisis de datos y el Big Data introducen o pueden introducir en diversos sectores de la sociedad.
- Adentrarse en el mundo de la Inteligencia de Negocio o Inteligencia Empresarial, comprendiendo en qué consiste y cuáles son sus características y sus principales ventajas.
- Conocer el programa que se empleará en el supuesto práctico, Power BI, así como las herramientas y/o componentes que lo forman, sus características y las ventajas que ofrece.

1.3. ESTRUCTURA DEL DOCUMENTO

En primer lugar, a modo de presentación, he escrito un prólogo. Seguidamente, aparece el resumen y las palabras clave, los cuales también vienen en inglés, por requerimiento de la Comisión de TFG, según especifica en el *Manual de Elaboración y Presentación de Trabajos de Fin de Grado*.

A continuación, aparecen los índices: el índice principal, el de siglas y acrónimos, el de figuras o ilustraciones y el de tablas.

El primer apartado es este, *Introducción y Objetivos*. En él se contextualiza y se justifica el trabajo y se exponen los objetivos del mismo. Además, se comenta la estructura del trabajo.

El segundo apartado, y el más grande de todo el bloque teórico, es el de *Análisis de Datos y Big Data*. En él, se estudia que son el análisis de datos y el Big Data, cuáles son sus características, ventajas y diferencias y cuál ha sido su evolución histórica. Se analizan algunos de los debates abiertos sobre estos conceptos y la posible evolución futura que pueden presentar. Además, se comentan bastantes ejemplos de sectores en los que el Big Data y el análisis de datos han mejorado las condiciones de trabajo, los resultados económicos y el marco social, entre otros muchos aspectos. En relación con el ámbito de este trabajo, también se estudia la inmersión de estas tecnologías en el ámbito deportivo, especialmente en el fútbol. Por último, se exponen una serie de conceptos asociados al análisis de datos y al Big Data, los cuales son fundamentales para conocer en profundidad estas ramas de estudio.



El tercer bloque es el de la *Inteligencia de Negocio*. Se estudia qué es este concepto, cuáles son sus características y sus principales ventajas. A mayores, se introducen los conceptos de minería de datos, como parte fundamental de la inteligencia empresarial, y los tableros de visualización, debido a la importancia que posteriormente adquirirán en el supuesto práctico.

El último apartado de la parte de teoría es *Power BI*. En este punto se comentan cuáles son las herramientas que componen el entorno Power BI, así como sus características principales y algunas ventajas. Para complementar, se presentan brevemente algunas herramientas alternativas que existen en la actualidad.

Para finalizar, se plantea, se desarrolla y se resuelve el *Caso Práctico*. En el apartado correspondiente se explica en qué consiste, cómo se ha desarrollado, qué resultados se han obtenido y cuáles son las conclusiones que se pueden extraer.

Lo último que aparece en el documento es la *Bibliografía* que se ha empleado. Dentro de este punto se debe a hacer una especial mención a la búsqueda e investigación que se ha realizado de otros TFG que hablan sobre los temas que aquí se tratan, con el objetivo de corroborar cuáles son los campos de estudio más interesantes desde un punto de vista técnico y tecnológico y de complementar los temas que ya se hayan tratado, dando una visión global a este trabajo.





2. ANÁLISIS DE DATOS Y BIG DATA

2.1. DEFINICIONES Y CARACTERÍSTICAS

Antes de adentrarse en el desarrollo y la exposición del trabajo, lo primero es describir el campo de conocimiento sobre el que vamos a trabajar. A continuación, vamos a ver qué es el análisis de datos y qué es el Big Data.

2.1.1. ¿Qué es el Big Data y cuáles son sus características?

Empezando por este último, algunas definiciones que podrían caracterizar al Big Data o datos masivos son las siguientes:

“Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles.

Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde 30-50 Terabytes a varios Petabytes.”

(PowerData, s. f.)

“El Big Data es el análisis masivo de datos. Una cuantía de datos, tan sumamente grande, que las aplicaciones de software de procesamiento de datos que tradicionalmente se venían usando no son capaces de capturar, tratar y poner en valor en un tiempo razonable.

Igualmente, el mismo término se refiere a las nuevas tecnologías que hacen posible el almacenamiento y procesamiento, además de al uso que se hace de la información obtenida a través de dichas tecnologías.”

(Facultad de Estudios Estadísticos UCM, s. f.)

Tras haber leído estas dos definiciones, ya podemos tener una idea más clara de lo que representa el término Big Data. Sin embargo, el concepto no se puede entender de forma completa sin conocer las características principales, las cuales son comúnmente conocidas como *las uves del big data*. Originalmente eran tres uves (volumen, velocidad y variedad), a las que posteriormente se añadieron otras dos (veracidad y valor). A día de hoy, algunos autores añaden otras dos uves adicionales (viabilidad y visualización). A continuación, se hará una breve exposición de cada una de ellas:

1.- Volumen: Es posiblemente la característica original del Big Data. La sociedad actual de la información genera cada minuto una cantidad tremenda de



datos, que necesitan ser tratados y analizados si no queremos perder una valiosa información y una importante fuente de negocios. Para hacernos una idea, en números, del volumen de datos que se genera cada minuto en el mundo digital, se muestra la siguiente ilustración:

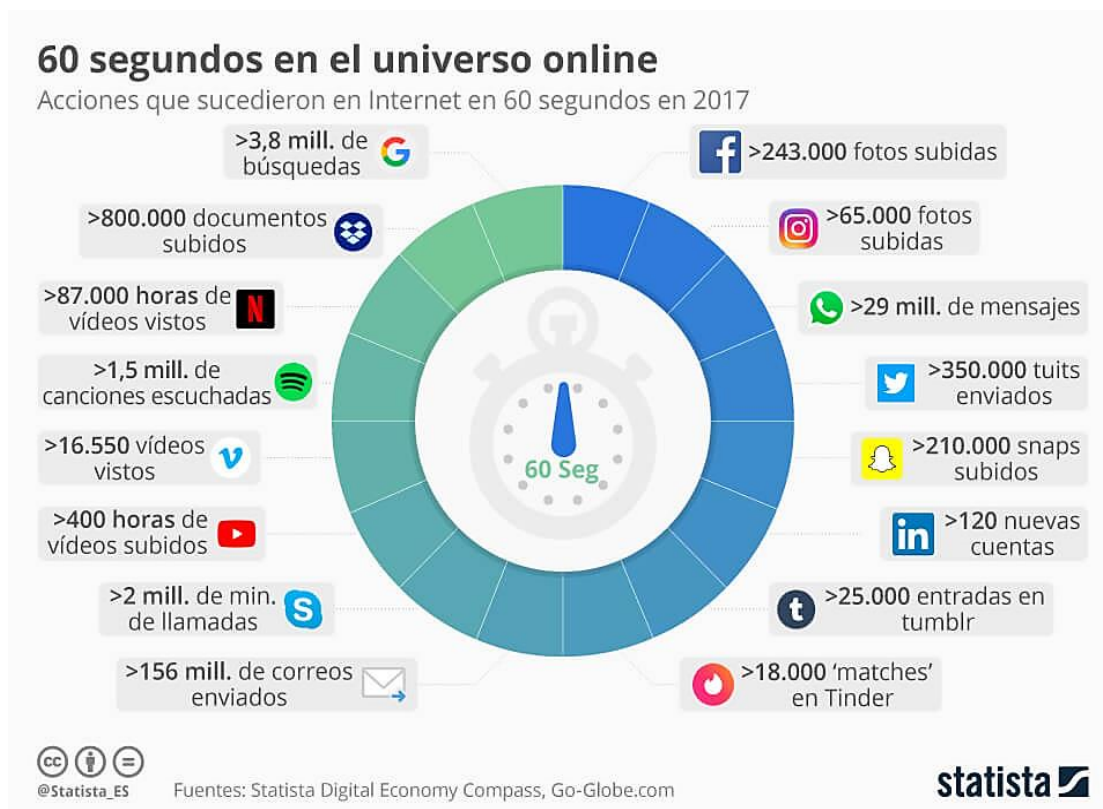


Ilustración 1. Datos generados en Internet en un minuto. Fuente: (Trabado, 2018)

2.- Velocidad:

“Además de manejar los datos, las empresas necesitan que esa información fluya con rapidez, lo más cercano posible a la inmediatez.”
(BBVA, 2017)

Si bien es cierto que, como se ha indicado antes, el volumen es la característica base del Big Data, también se debe recalcar que es la velocidad la que generalmente otorga la ventaja competitiva a las empresas. En otras palabras, son aquellas compañías y organizaciones que son capaces de tratar los datos de una forma más rápida las que sacan mayor provecho del uso del Big Data y de sus herramientas asociadas.

“Los datos tienen que estar disponibles en el momento justo para tomar las decisiones adecuadas de negocio.”
(BBVA, 2017)



3.- Variedad:

“Es fundamental hacer frente a la variedad de datos que aumenta el grado de complejidad, tanto en el almacenamiento de información como en el análisis. Por ello, hay que dar uniformidad a los datos que tienen origen heterogéneo, porque una de las fortalezas del Big Data es que se puede conjugar y combinar todo tipo de información para alcanzar un todo homogéneo.”

(Big Data International Campus, 2020)

Los datos pueden ser estructurados, semiestructurados o no estructurados. Los primeros son los que tendríamos en una base de datos clásica, con un formato y esquema bien definidos. Por su parte, los datos no estructurados son aquellos que proceden de diversas fuentes y que se presentan en múltiples formas sin seguir un formato común. Son los que más problemas generan tanto en la obtención como en el tratamiento. Por último, los semiestructurados son datos que no pueden ser considerados como estructurados porque presentan algunas diferencias, pero que sí que presentan un relativo grado de uniformidad, que facilita su tratamiento.

4.- Veracidad:

“La cuarta uve es de veracidad, que en este contexto equivale a calidad. Tenemos todos los datos, pero ¿se nos puede escapar algo? ¿Son datos ‘limpios’, veraces? ¿Realmente aportan algo?”

(BBVA, 2017)

Esta característica depende directamente del tratamiento que se le dé al conjunto de datos objeto de estudio. Son necesarias una serie de herramientas que “limpien” los datos obtenidos, para lograr así un óptimo aprovechamiento de estos. Cuanto mejores sean los datos, más posibilidades hay de que las decisiones tomadas sean las adecuadas.

5.- Valor:

“Por último, en la cúspide de la pirámide del Big Data reina la uve de valor, es decir, la capacidad de transformar todo ese tsunami de datos en negocio.”

(BBVA, 2017)

Resulta obligado indicar que esta dimensión afecta, en mayor o menor medida, a todas las demás. Está comúnmente aceptado que este factor es el más relevante del Big Data.

“En definitiva, la clave está en cómo obtener la mejor información, el mejor valor y conocimiento para sacar la mayor rentabilidad.”

(Big Data International Campus, 2020)



Hasta aquí llegan las 5 uves del Big Data. Sin embargo, como ya he indicado antes, actualmente hay autores que suman dos uves más a la lista. Son la viabilidad y la visualización de los datos. A continuación, veremos en qué consiste cada una:

6.- Viabilidad: No basta con conseguir los datos, tratarlos y extraer conclusiones, sino que también es necesario realizar todo este proceso de una forma eficiente. En caso contrario, la tecnología Big Data se volvería inviable para la gran mayoría de empresas y corporaciones.

7.- Visualización: Normalmente, las personas encargadas de tomar las decisiones no son las que han estado trabajando con los datos, por lo tanto, se las debe informar convenientemente de las conclusiones obtenidas. Además, tal y como ya se ha comentado, el tiempo suele jugar un papel importante, lo que obliga a que la información transmitida esté bien resumida y que sea clara y concisa, pero sin perder información valiosa por el camino.

En este contexto, se antoja imprescindible una correcta visualización de los datos. El uso de gráficos, esquemas, ilustraciones, paneles dinámicos... es cada día más frecuente e importante. En este trabajo también cobrará especial relevancia este aspecto y, como se verá en su apartado correspondiente, el programa Power BI nos permitirá obtener una más que satisfactoria visualización de los datos.

En la siguiente ilustración, se puede ver un resumen de *las 7 uves del Big Data* y cómo están interrelacionadas en la cadena de creación de valor:



LAS 7 Vs del BIG DATA

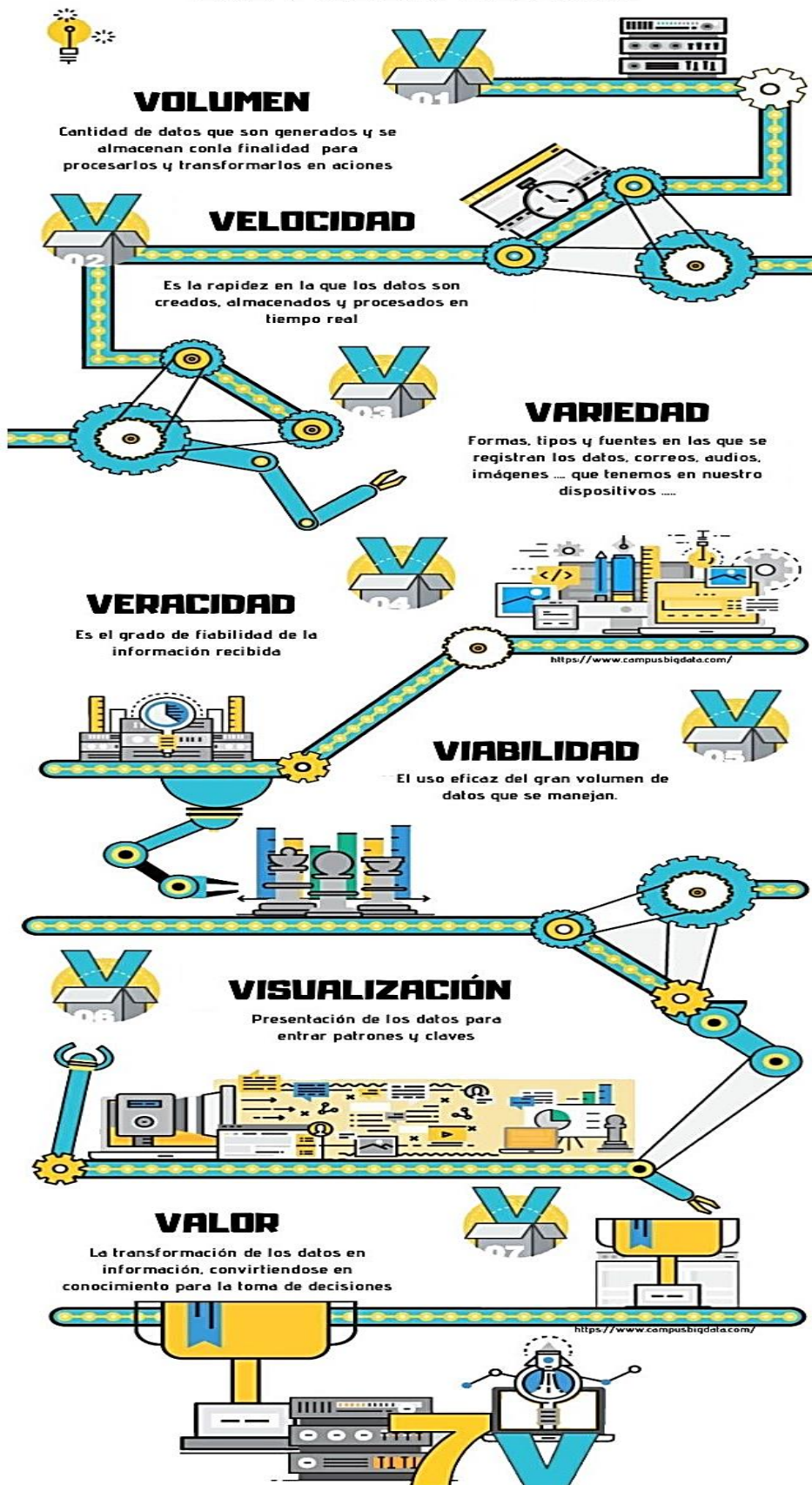


Ilustración 2. Las 7 uves del Big Data. Fuente: (Big Data International Campus, 2020)



2.1.2. ¿Qué es el análisis de datos y cuáles son sus características, objetivos y ventajas?

En lo que respecta al ámbito de este trabajo, la diferencia que justifica porqué el caso práctico que más adelante elaboraré se engloba dentro del análisis de datos y no dentro del Big Data es el volumen o tamaño del conjunto de datos con los que trabajaré. Tal y como aparece reflejado al comienzo del subapartado anterior, el tamaño aceptado para determinar si un conjunto de datos se considera Big Data va desde entre 30 y 50 Terabytes a varios Petabytes. El volumen de datos que emplearé se encuentra muy alejado de esas cifras.

Sin embargo, he considerado incluir en la parte teórica apartados tanto de Big Data como de análisis de datos ya que, desde mi punto de vista, el primero presenta un mayor interés tecnológico y es una rama de estudio más actual y con aún muchas utilidades y aplicaciones por descubrir. Pero sí que se hace imprescindible resaltar que el Big Data y el análisis de datos no son, ni mucho menos, ramas contrarias, sino que son complementarias. En otras palabras, se puede afirmar que el Big Data ha ampliado los horizontes de actuación del análisis de datos.

Centrándonos de nuevo en el análisis de datos, algunas definiciones que nos pongan en contexto pueden ser las siguientes:

“El análisis de datos es un proceso que consiste en inspeccionar, limpiar y transformar datos con el objetivo de resaltar información útil, para sugerir conclusiones y apoyo en la toma de decisiones.”

(Wikipedia, 2020a)

“El análisis de datos es la ciencia que se encarga de examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información para poder tomar decisiones, o simplemente ampliar los conocimientos sobre diversos temas.”

(QuestionPro, 2018)

“El análisis de datos se refiere a las técnicas y procesos cualitativos y cuantitativos utilizados para mejorar la productividad y la ganancia de los negocios. Los datos se extraen y categorizan para identificar y analizar datos y patrones de comportamiento, y las técnicas varían según los requisitos de la organización.”

(tecnologias-informacion.com, 2018)

Tras haber leído estas definiciones del análisis de datos, vamos a exponer cuáles son los tipos de datos que existen, algo que es necesario conocer antes de realizar cualquier trabajo o proyecto de las características de este TFG. Los podemos ver en la siguiente tabla:



	Tipos de datos	Análisis	Ejemplos
Cualitativo	Se centra en las opiniones, actitudes y creencias.	Preguntas y respuestas a preguntas como: ¿Por qué? ¿Cómo?	Paneles en donde se da una discusión y se entrevista a consumidores sobre lo que les agrada o no del lugar.
Cuantitativo	Se centra en los datos duros e información que pueda contabilizarse.	Se obtiene mediante preguntas similares a: ¿Cuántos? ¿Quién? ¿Con qué frecuencia? ¿Dónde?	Encuestas enfocadas a medir las ventas, tendencias, reportes o percepciones.

Tabla 1. Tipos de datos. Fuente: (QuestionPro, 2018)

Desde un punto de vista técnico, los datos cuantitativos son los que presentan un mayor interés y una relación más directa con el campo de la ingeniería en organización industrial. Por ello, es a los que se dedicará la mayor parte de este trabajo. Sin embargo, no es menos cierto que los datos cualitativos también pueden, en función de las circunstancias, marcar la diferencia en la toma de decisiones.

Un claro ejemplo de esto es el ámbito de la gestión deportiva. A la hora de fichar a un jugador, los datos cuantitativos como velocidad punta, goles por partido, pases acertados... son fundamentales y muestran un amplio abanico de las características del jugador. Pero son aspectos como el entorno familiar, la evolución académica o las habilidades sociales los que en multitud de ocasiones inclinan la balanza por uno u otro jugador. La imposibilidad de cuantificar estos indicadores es una dificultad añadida que los ojeadores y directores deportivos deben sortear. En posteriores apartados se profundizará más en este tema.

En este punto, ya conocemos algunas posibles definiciones del análisis de datos y qué tipos de datos nos podemos encontrar. Con ello, lo siguiente que vamos a estudiar son los objetivos generales que se pretenden conseguir al llevar a cabo un análisis de datos. Son los siguientes:

- “1.- Resumir el conjunto de variables en unas pocas nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
- 2.- Encontrar grupos en los datos si existen.
- 3.- Clasificar nuevas observaciones en grupos definidos.
- 4.- Relacionar dos conjuntos de variables.”

(Peña, 2002)

Lógicamente, cada análisis particular presentará a mayores una serie de objetivos específicos que dependerán de multitud de factores. Para el caso práctico que expondré, se estudiarán en detalle cuáles son esas metas específicas y que herramientas se pueden emplear para abordarlas.

Volviendo a los objetivos generales, podemos encontrar varias relaciones entre estos y las siete *aves del Big Data* que vimos en el apartado anterior. Por



ejemplo, reducir el número de variables está relacionado con el volumen, pues en función de las herramientas informáticas y tecnológicas de las que dispongamos podemos tener problemas a la hora de manejar muchas variables al mismo tiempo. También guarda relación con la velocidad, pues es posible que un estudio exhaustivo de muchas variables llevara un tiempo inasumible. A veces, es mejor perder un poco de información a cambio de obtener resultados rápidamente.

Por su parte, los objetivos tres y cuatro se encuentran ligados al valor. El hecho de poder clasificar nuevas observaciones y relacionar dos conjuntos de variables aporta un valiosísimo conocimiento a la empresa, el cual puede ser fuente de innumerables ventajas competitivas.

Lo último que se tratará en este subapartado son las ventajas que, a nivel empresarial, nos puede ofrecer un correcto análisis de datos:

- *“Capacidad para tomar decisiones de negocios más rápidas e informadas, respaldadas por hechos.*
- *Ayuda a las empresas a identificar problemas de rendimiento que requieren algún tipo de acción.*
- *Comprensión más profunda de los requisitos de los clientes, lo que, a su vez, crea mejores relaciones comerciales.*
- *Mayor conciencia del riesgo, permitiendo la implementación de medidas preventivas.*
- *Puede verse de forma visual, lo que permite tomar decisiones más rápidas y mejores.*
- *Puede proporcionar a una empresa una ventaja sobre sus competidores.*
- *Mejor conocimiento del desempeño financiero del negocio.*
- *Se ha demostrado que reduce los costos y, por lo tanto, aumenta los beneficios.”*

(QuestionPro, 2018)

Sin embargo, todas estas ventajas no se lograrán si el análisis de datos no se hace de un modo ordenado y estructurado. Para ello, lo primero que tenemos que hacer es definir cuál es el problema al que se ha de dar solución, pues, en caso contrario, es fácil acabar yéndose por las ramas. Seguidamente, se deben establecer los indicadores que se utilizarán para medir y evaluar el grado de consecución de los objetivos. Además, es necesario precisar cómo se van a medir dichos indicadores. En tercer lugar, está la obtención o captación de los datos. Esta es una etapa muy importante, por lo que debemos conocer con precisión las fuentes y bases de datos con las que vamos a trabajar. Es imprescindible disponer de un sistema de almacenamiento y establecer unos registros que mantengan los datos organizados. El cuarto paso es el propio análisis. Filtrar información, encontrar tendencias, ordenar y clasificar archivos, etc. son actividades laboriosas que requieren concentración y unos conocimientos avanzados por parte del analista. En último lugar, se encuentran la interpretación de resultados y la obtención de conclusiones.



2.2. HISTORIA DEL ANÁLISIS DE DATOS Y EL BIG DATA

2.2.1. Origen, evolución y situación actual

Algunos pensarán, especialmente cuando se habla del término *Big Data*, que este conjunto de conceptos ha surgido hace relativamente poco tiempo. Muchos lo relacionarán directamente con la aparición de los ordenadores y la informática, en el siglo XX. Sin embargo, nada más lejos de la realidad.

Si nos dedicáramos a buscar e investigar sobre los primeros antecedentes de registros de información con un fin de predicción, veríamos que se remontan a la época del Paleolítico, tal y como menciona (BusinessGoOn, 2019). Se ha demostrado que las tribus hacían muescas en huesos y piedras para llevar un registro o control de las actividades diarias y comerciales, con el fin de gestionar la duración del suministro de alimento y de otros bienes. Años más tarde, han tenido lugar varios descubrimientos de la humanidad que han reforzado el interés por los datos, como la invención del ábaco que, según indica (BusinessGoOn, 2019), supuso un empuje determinante al cálculo y análisis. También se debe hablar de la aparición y consolidación de las primeras grandes bibliotecas, como la de Alejandría, que marcan un antes y un después en la raíz del almacenaje y tratamiento de datos.

(BusinessGoOn, 2019) también menciona que, a finales del siglo XVII, John Graunt llevó a cabo el primer ensayo registrado de análisis de datos estadísticos sobre defunciones, con el cual logró diseñar un sistema de alerta temprana de la terrible peste bubónica, que durante ese siglo assolaba Europa. Dos siglos más tarde, en 1865, Richard Millar Devens empleó por primera vez el término “business intelligence” en su *Enciclopedia de Anécdotas Comerciales y de Negocios*, donde narra cómo el banquero y comercial Henry Furnese logró obtener algunas ventajas sobre sus competidores al recopilar, tratar y analizar información trascendental de sus actividades y labores comerciales.”

En el año 1880, “Herman Hollerith, empleado del censo estadounidense, desarrolla su máquina tabuladora. Con ella consigue reducir un trabajo de 10 años a 3 meses. Este ingeniero funda una compañía que posteriormente se conocería como IBM. En 1928, el ingeniero alemán Fritz Pflueumer patenta el primer sistema magnético para almacenar datos. Sus principios de funcionamiento se utilizan hoy en día.”

(Grupo IGN, 2017)

En torno a los años 30 y 50 del siglo XX comenzó la era de la computación. (ARISER, 2015) nos indica que, en el año 1940, Turing y Good diseñaron una máquina capaz de analizar documentos cifrados por los alemanes usando la máquina enigma en el transcurso de la Segunda Guerra Mundial, de tal manera que era capaz de descodificar la clave con la que habían sido cifrados. (ARISER, 2015) también relata que, en ese mismo año, se desarrolla una máquina, el Predictor de Kerrison, preparada para automatizar la defensa antiaérea, apuntando contra aviones enemigos. Además, en el año 1946, el equipo *Proyecto Manhattan* consigue



utilizar ordenadores para analizar y predecir el comportamiento que podría causar una reacción nuclear en cadena.

Otro hecho destacable tiene lugar a lo largo de 1956. En ese año se consigue resolver el problema del camino más corto mediante el empleo de la analítica computacional, lo que permite obtener importantes mejoras en los campos de la logística y el transporte. En 1958, tal y como viene relatado en el artículo de (ARISER, 2015), FICO implementa modelos predictivos para las decisiones en los riesgos de créditos. Posteriormente, en 1966, comienza lo que más adelante se conocerá como SAS Institute, como un proyecto de investigación en el área de la analítica, el cual estaba financiado por el Ministerio de Agricultura de los Estados Unidos.

En 1958, “el informático alemán Hans Peter Luhn, define la inteligencia de negocio: la habilidad de percibir las interrelaciones de los hechos presentados para guiar acciones hacia un objetivo deseado.”

(Grupo IGN, 2017)

Según avanza el siglo XX, la creciente generación de datos es cada vez más patente, por lo que aumenta exponencialmente la necesidad de poder registrarlos y analizarlos a tiempo. Tal y como cuenta el artículo de (BusinessGoOn, 2019), en el año 1965, el gobierno de los Estados Unidos concibió el primer centro de procesamiento de datos del mundo, en el cual se podía almacenar en cinta magnética más de 740 millones de declaraciones de impuestos y en torno a 175 millones de huellas dactilares.

“Entre 1970 y 1999 el análisis se populariza, creándose en 1973, por ejemplo, el modelo Black-Scholes para predecir el precio óptimo de las acciones en el futuro, o la primera herramienta comercializada para construir sistemas de apoyo en las decisiones guiadas por modelos en 1980. Pero no es hasta bien entrada la época de los 90, en 1995, cuando se lanzan las webs Amazon y eBay, que suponen el comienzo de la carrera para la personalización de las diversas experiencias online para cada uno de los usuarios. Esto supone la necesidad de que las búsquedas de los motores tengan en cuenta cada vez más la importancia de la relevancia de los resultados, llegando a aplicarse en 1998 algoritmos de búsqueda en Google.”

(ARISER, 2015)

Los avances en el ámbito del tratamiento de datos se van multiplicando a lo largo de la segunda mitad del siglo XX, y es en el último cuarto del siglo cuando aparecen algunos de los elementos que, a la postre, cambiarían el mundo para siempre. Por ejemplo, como se puede leer a continuación, la aparición de Internet:

En 1970, “IBM desarrolla el modelo relacional de base de datos, gracias al matemático Edgar F. Codd. En 1976, se populariza el uso de MRP (software de gestión de materiales), antecedentes de los ERP actuales, que mejoran la



eficiencia de las operaciones en la empresa; además de generar, almacenar y distribuir datos en toda la organización.

En 1989, Erik Larson habla por primera vez de Big Data en el sentido que conocemos la expresión hoy en día. La revista Harpers Magazine recoge su artículo, en el que especula sobre el origen del correo basura que recibe. En torno a este año se empiezan a popularizar las herramientas de business intelligence para analizar la actividad comercial y el rendimiento de las operaciones. En 1991 nace Internet, a la postre, la gran revolución de la recolección, almacenamiento y análisis de datos. Tim Berners-Lee establece las especificaciones de un sistema de red con interconexiones a nivel mundial accesible para todos en cualquier lugar.”

(Grupo IGN, 2017)

Hacia finales del siglo, los costes del almacenamiento de datos comienzan a ser accesibles, lo que impulsa definitivamente al Big Data. Tal y como se nombra en el artículo de (Grupo IGN, 2017), el libro *La Evolución de los Sistemas de Almacenamiento*, del año 2003, establece esta fecha como el primer año de la historia en el que el almacenamiento en formato digital es más barato que el papel.

La última fecha en la que me detendré es 1997, año en el que Google lanza su sistema de búsqueda en Internet, el cual lidera, a nivel mundial, el número de búsquedas en la red, hasta la actualidad.

Hasta aquí se ha hablado del origen y la evolución del análisis de datos y del Big Data. Como se ha explicado, aunque el origen de lo que entendemos hoy en día al hablar de estos conceptos se sitúa en el siglo XX, el interés del ser humano por obtener, procesar e interpretar información se remonta a la prehistoria.

Cabe destacar que, aunque parece que es el mundo empresarial y de negocios el que más partido saca de estas tecnologías, su desarrollo y evolución guarda una interesante relación con la industria militar y la guerra. Por último, se debe valorar el giro que, desde el último cuarto del siglo XX, se viene produciendo. Y este no es otro que el carácter social del Big Data y el análisis de datos. Monitorizar los comportamientos, movimientos, actitudes e incluso pensamientos de las personas se ha convertido en la obsesión de gobiernos y empresas en todo el mundo. Los primeros para controlar y gestionar a la población y los segundos para comerciar y obtener un rédito económico.

Tras haber desarrollado estas consideraciones, acabaré este subapartado valorando la situación actual.

En primer lugar, se ha de explicar que el Big Data es un concepto cada vez más generalizado y que cada día más empresas desean y necesitan disponer de sus herramientas.



“Varios factores explican la creciente adopción de las tecnologías Big Data en los últimos años. Por un lado, tenemos una sensible reducción en la barrera económica de entrada a esta tecnología, gracias al amplio conjunto de herramientas de código abierto que implementan los avances en este campo, y a la disponibilidad cada vez más asequible a plataformas de Cloud Computing (computación en la nube). [...] A esto se une también el auge de conceptos como Internet of Things (Internet de las cosas), que permiten la proliferación de todo tipo de dispositivos capturando y compartiendo datos para analizar diversos procesos en diferentes campos de aplicación. Todo lo anterior, junto a la enorme repercusión mediática que adquieren muchas de las innovaciones tecnológicas generadas en entornos como Silicon Valley, explica la popularización del término Big Data y el interés por parte del público en general, más allá de contextos especializados.”

(NIÑO et al., 2015)

En segundo lugar, mencionaré algunos conceptos como el aprendizaje automático, el cual constituye de uno de los principales potenciales que presenta el estudio de los datos. Esta es aún una tecnología que está dando sus primeros pasos, pero todo indica que en un futuro próximo se producirán avances muy importantes. Siguiendo en la misma línea, se puede hablar de la inteligencia artificial. Profundizaré sobre estos y otros conceptos en apartados posteriores.

Por último, siendo la razón de ser de este trabajo, he de indicar que las tecnologías asociadas al análisis de datos y al Big Data cada día se expanden a más campos y áreas. La medicina es una de las ciencias que más se puede aprovechar y, posiblemente, veremos como la tecnología ayuda a salvar miles de vidas y a curar cientos de enfermedades en los próximos años. El deporte y su afán por la mejora continua también han visto como un correcto análisis de datos y una adecuada interpretación de la información pueden ser determinantes para lograr los objetivos marcados. La política, a través de, por ejemplo, las campañas electorales, es otro ejemplo mencionable.

2.2.2. El futuro. Posible evolución y debates abiertos

Un trabajo de estas características carecería totalmente de sentido si no se hace un breve análisis y una valoración de la evolución futura, la utilidad y las aplicaciones de aquello sobre lo que se está hablando y, en este caso, con mucha más razón aún. Como bien sabemos, el mundo del Big Data y del análisis de datos se encuentra en pleno desarrollo y evolución y, a la par que la informática y la tecnología, las mejoras en estos campos hacen que las utilidades y aplicaciones aumenten exponencialmente año a año.

Todo parece indicar que, en el futuro, todos los sectores imaginables podrán sacar provecho de los datos y la información. Algunos de los más destacados son:

La medicina, como ya se ha mencionado anteriormente y en el cual profundizaré en próximos apartados, donde la posibilidad de contar con un registro



de muchísimas variables como constantes vitales, índices corporales, estado de las células y de determinados órganos... puede ayudar a la detección precoz de innumerables patologías y enfermedades, así como a acelerar y a mejorar los procesos de recuperación, los seguimientos posteriores, etc.

La agricultura, donde se puede optimizar el uso del suelo, del agua, de los abonos y de la maquinaria. Además, se podría monitorizar el estado de las cosechas, para saber en tiempo real su evolución. Otro campo fundamental es el de la educación, especialmente cuando se refiere a la educación a distancia o a la introducción de recursos digitales en la presencial.

Otros campos en los que el Big Data ya se ha adentrado, pero que sin duda evolucionarán notablemente en los próximos años, son la publicidad (publicidad personalizada en Internet, optimización empresarial de las campañas...), la administración pública (trámites electrónicos, control de censos, gestiones burocráticas, mayor rapidez en la Justicia, estudios estadísticos, Inteligencia y cooperación internacional, etc.), innumerables áreas de la ingeniería (gestión, control de plantas, mantenimiento preventivo y predictivo, construcción, etc.) y, por supuesto, la gestión deportiva.

Todo lo que se ha comentado hasta ahora muestra la cara positiva y útil de la introducción de estas tecnologías en el día a día de la sociedad. Sin embargo, no todo es positivo, y existen multitud de debates abiertos a los que se deberá dar respuesta en los próximos años. La privacidad y la seguridad lideran la mayoría de estos interrogantes.

“Hay voces que ya se alzan y advierten sobre el peligro de que gobiernos y empresas como Google, Apple y Facebook, entre otras, tengan en su poder información tan sensible como nuestras fotos y correos electrónicos, puedan monitorear nuestros movimientos físicos en tiempo real, saber la cantidad de dinero que tenemos en el banco, cuándo y dónde compramos la comida, qué serie de televisión vemos, a dónde nos iremos de viaje y a qué hotel, los cumpleaños de nuestros hijos, conocer nuestros deseos, lo que buscamos en la Web y lo que está ocurriendo en tiempo real en la intimidad de nuestro hogar. Perdón, ¿dije intimidad? ¿Todavía alguien cree que existe la privacidad?”

(Zanoni, 2014)

Evidentemente, si bien es cierto que las redes sociales y los teléfonos inteligentes han facilitado las relaciones entre las personas y la posibilidad de comunicarnos con gente a muchísima distancia, no es menos cierto que esto ha supuesto disminuir considerablemente nuestra privacidad personal y nuestra intimidad en el día a día. Y, por supuesto, todo ese “rastros” y esa información que voluntariamente dejamos en el mundo digital es utilizado por empresas y organizaciones con multitud de fines, no todos positivos.



Es también importante resaltar que la evolución de la tecnología y de la ciencia va directamente ligada a la evolución de la sociedad y de la forma y hábitos de vida.

“Cada vez seremos más personas en el mundo. Y en pocos años, exactamente al revés de cómo era hace cien años, el 80 por ciento del planeta vivirá dentro de las grandes ciudades. [...] Las megalópolis tienen y tendrán más urgencias y problemas de compleja resolución: seguridad, movilidad y transporte, planeamiento urbano, servicios públicos, higiene, conectividad, desarrollo sustentable, salud, recursos energéticos, vivienda, alimentación, comunicación, etc. Sin dudas, grandes desafíos para los gobiernos, las empresas y universidades, entre otras instituciones.”

(Zanoni, 2014)

Estas cuestiones son algunas de las que se pretenden abarcar con el Internet de las Cosas, tema sobre el que se profundizará en apartados posteriores, el cual no podrá triunfar sin una correcta aplicación del análisis de datos y de tecnologías asociadas al Big Data.

El último aspecto que quiero valorar en este punto es el de la seguridad. El mundo digital es un nuevo campo de actuación para multitud de delincuentes. Muchos países se están viendo en la obligación de aprobar leyes y normativas al respecto, ante la proliferación de nuevos delitos informáticos. Términos como suplantación de identidad, ciberacoso, robo de contraseñas, virus informáticos... estarán cada día más presentes en nuestra sociedad.

“Para conseguir una respuesta eficaz a los peligros que amenazan las nuevas redes de comunicación, es sumamente importante y prioritario identificar las vulnerabilidades e individualizar los peligros existentes y potenciales que dichas vulnerabilidades permiten. Esto se consigue con un trabajo de inteligencia, o mejor dicho de ciber inteligencia. [...] La investigación en Internet tiene otra dificultad añadida que afecta a todas las policías del mundo: el espacio de Internet carece de fronteras y el contenido ilícito –como la pornografía infantil– circula de un país a otro en milésimas de segundos. [...] Cada vez más se emplean técnicas de comunicaciones a través de Internet para cometer uno de los delitos más tradicionales de la sociedad: el del engaño para conseguir un beneficio económico.”

(Carlos & Castilla, 2006)

2.2.3. Historia del Análisis de Datos y el Big Data en el fútbol

Tal y como ya he indicado en algunos puntos, el fútbol es el marco que he escogido para desarrollar este trabajo por su carácter global, comúnmente conocido y atractivo. Sin embargo, este es un TFG de Ingeniería en Organización Industrial, por lo que la mayor parte de la teoría se ha centrado (y se centrará) en las cuestiones técnicas y tecnológicas, que guardan mayor relación con los conocimientos adquiridos a lo largo del grado.



No obstante lo anterior, he creído oportuno terminar el apartado de historia con algunas breves reseñas sobre la evolución del análisis de datos y del Big Data en el mundo del fútbol.

Desde sus orígenes, el fútbol no ha dejado de evolucionar y cambiar permanentemente. Año tras año vemos como, estilos de juego que habían traído éxitos y victorias a aquellos que los habían practicado, dejan de generar buenos resultados y obligan a modificarlos por completo.

Ese progreso embellece a este deporte y genera un mayor atractivo. Lógicamente, si siempre triunfaran los mismos estilos de juego, el fútbol acabaría perdiendo interés, pues siempre habría que hacer lo mismo para ganar y siempre se sabría quien iba a vencer: el que mejor hiciera eso.

Tal y como destaca (Antón Carranza, 2017), han sido muchos los que han buscado evolucionar el fútbol y la forma de jugar o crear un método que resultara exitoso y vencedor. Algunos prefirieron buscar el desarrollo de mejoras técnicas, otros, implementaron nuevos sistemas y métodos de entrenamiento físico. También hubo quien decidió apostar por una u otra táctica determinada; y hubo algunos que pretendieron apoyarse en las matemáticas y la estadística como fuente de generación de ventajas competitivas y de conocimiento deportivo. En otras palabras, algunos pretendieron ir más allá en lo que a análisis en la gestión deportiva se refiere. Anhelaron conocer y desmenuzar el funcionamiento del fútbol, constatar qué es lo que encierra el juego y recopilar todos estos hechos y acciones en forma de datos, para obtener una información a través de la cual extraer conclusiones y, por consiguiente, deducir conocimiento.

(Antón Carranza, 2017) resalta que uno de los primeros analistas del fútbol, por no decir el primero, fue el inglés Charles Reep, que comprendió que el estudio y análisis de acciones pasadas podía ser provechoso para predecir eventos futuros y que, implementado un registro de datos (lo que hoy se conocería como una base de datos) con estos hechos, sería verosímil analizar los partidos de fútbol. Charles Reep fue una de las primeras personas que trató de demostrar con datos que no todo lo que nuestro conocimiento y/o experiencia retiene, ni aquello que nuestra impresión nos expone, es absoluto o incuestionable en el fútbol.

“En 1967, presentó junto con Bernard Benjamin un informe donde revelaba determinados patrones de comportamiento asociados a los eventos de cada partido. Tras el análisis recopilado durante 15 años determinaron que:

- En promedio, los equipos marcaron gol en aproximadamente uno de cada nueve intentos.

- Las probabilidades de realizar un pase acertado por cada equipo eran cercanas al 50%, pero dicha probabilidad iba disminuyendo conforme se incrementaba el número de pases. Es decir, a más pases, menos probabilidad de que el siguiente llegue al destino deseado.



- El 30% de todas las recuperaciones efectuadas en el área rival condujeron a tiros a portería, y casi la mitad de todos los goles procedían de esas recuperaciones.

Pero sin duda, la afirmación más contundente e influyente, fue aquella en la que aseguró que la mayoría de los goles procedían de la secuencia de 3 o menos pases. En su opinión, se debía tratar de minimizar los esfuerzos (números de pases) para maximizar los objetivos (número de goles). La clave para ganar partidos parecía ser tan simple como reducir el número de pases y la posesión, tratando de que el balón llegara lo más rápido posible al área rival. Una forma de conseguir dicho propósito es mediante el juego directo.”

(Antón Carranza, 2017)

Posiblemente Charles Reep estuviera equivocado en muchos de sus postulados, pero el éxito que sus ideas tuvieron en bastantes equipos y durante muchas temporadas (incluso algunas ideas serían válidas en la actualidad), hacen que él fuera el primero en demostrar que merece la pena analizar, desde un punto de vista estadístico, qué es lo que sucede en el terreno de juego.

Según expone (Hereña, 2018), las técnicas pioneras de registro de información, en el ámbito del fútbol, fueron denominadas “lápiz y papel”, llevadas a cabo por el citado Charles Reep. Consistían en anotar o apuntar en un papel los movimientos, desplazamientos y las acciones de los jugadores. Estos métodos, a pesar de ser sencillos y económicos, consumían demasiado tiempo y su validez dependía también de otros factores. Las líneas de desarrollo e investigación fueron evolucionando, hasta lograr desarrollar una nueva técnica de registro: las grabaciones magnetofónicas. Estos sistemas, por medio de la verbalización, examinaban los movimientos y desplazamientos de los futbolistas, obteniendo la distancia recorrida al cabo de un partido.

“La siguiente evolución llegó de la mano de los sistemas de filmación y permitió mejorar las técnicas precedentes debido, principalmente, a que se podían analizar los partidos en diferido. Esto permitió ampliar el número de los indicadores de rendimiento que se podían investigar, convirtiendo la observación a través del vídeo en una técnica muy utilizada. La incorporación de la informática fue otro paso importante, ya que permitió una reducción del porcentaje de error, además de suponer la aparición de nuevas técnicas de análisis que permitían registrar más información. [...] Al final de los años noventa, el análisis de los movimientos en el fútbol sufrió una revolución con el desarrollo comercial de los sistemas de seguimiento semiautomáticos. Estos requieren de la instalación de múltiples cámaras alrededor del terreno de juego, además de sistemas informáticos muy específicos y costosos.”

(Hereña, 2018)

Para concluir este apartado, creo importante destacar algo que ya se ha comentado en apartados anteriores y es que, pese a la introducción de la tecnología



en el deporte, la figura del ojeador u observador sigue siendo igual de importante. Esto es así porque los datos aún son incapaces de mostrar por sí solos aspectos como el entorno del jugador, su respuesta ante determinadas situaciones adversas, su conducta social, etc. por lo que, por ahora, ese trabajo debe ser realizado por personas.

2.3. CAMPOS DE APLICACIÓN: ALGUNOS EJEMPLOS

2.3.1. Medicina

He insistido bastante en algunos de los apartados anteriores, y ahora lo pongo en primer lugar, en el campo de la medicina, porque considero que es uno de los más apasionantes y uno en los que se pueden lograr más y más importantes cosas con la aplicación de herramientas de Big Data y de Análisis de Datos.

“En Internet, la información sobre las enfermedades y sus brotes se difunde no sólo a través de noticias de las agencias de los gobiernos, sino también por canales informales, que van desde la prensa a los blogs, mediante análisis de los registros de búsquedas en la web. [...] Una publicación pionera se centró en la localización de registros de consultas de Google para detectar la actividad de la gripe en regiones específicas con grandes poblaciones de usuarios que hacen búsquedas en las webs. [...] Por otra parte, los big data proporcionan un complemento útil en la medida en que se han convertido en un componente importante para la vigilancia de enfermedades infecciosas como la gripe.”

(Parra Calderón, 2016)

Ya había comentado algunas cuestiones sobre la vigilancia y el seguimiento de enfermedades e infecciones. Sin embargo, como vemos, la aplicación de la tecnología del Big Data en la medicina va mucho más allá. Controlar qué enfermedades son las que más búsquedas reciben en Internet puede servir para identificar la “ruta” que sigue un virus, algo fundamental como estamos comprobando en la crisis del coronavirus COVID-19.

Los distintos planteamientos del Big Data permiten integrar la capacidad de geolocalización, obtenida por la dirección de los ciudadanos, así como por la dirección de establecimientos tales como las farmacias, los hospitales y los ambulatorios, lo que, tal y como menciona (Parra Calderón, 2016), hace viable elaborar mapas geográficos de salud en el tiempo para toda la población de una localidad, una región o incluso un país. Una aplicación de futuro que puede afrontar el Big Data, en materia de salud pública, es la incorporación e integración de otras fuentes de información, como son los contaminantes, el tráfico urbano, las tiendas de comestibles y los mercados, entre otras fuentes, que podrían mejorar la concreción de la estratificación por peligro o riesgo de la población objeto de análisis.

“Las aplicaciones de Big Data en el sector de la salud indican un alto potencial para mejorar la eficiencia y calidad de provisión de cuidados. En el



estudio de McKinsey se afirma que, si el sector de la salud en los Estados Unidos usara tecnologías de Big Data de manera efectiva para producir calidad, el sector podría crear más de 300 billones de dólares cada año y dos tercios de esto sería en forma de reducción de gastos de salud en alrededor de un 8%. [...] En concreto en España, el gasto en Tecnologías de Información y Comunicaciones (TIC) en la sanidad pública en 2013 ascendió a 624 millones de euros, lo que significa que menos del 1,20% del gasto sanitario público total se dedicó a las TIC.”

(Menasalvas, Ernestina; Gonzalo, Consuelo; Rodríguez-González, s. f.)

Como cualquier otro sector, la medicina también debe tener en cuenta el apartado económico y, tal y como podemos leer, también ahí el uso de tecnologías Big Data ofrece innumerables ventajas.

“Del estudio se desprende que, en la fecha del estudio, las TIC se estaban utilizando para almacenar la información generada por el sistema sanitario de salud, pero no se estaba utilizando dicha información para mejorar la atención a los pacientes y la gestión y eficiencia de los servicios sanitarios. La Historia Clínica Digital del Sistema Nacional de Salud (HCDSNS) surge para garantizar a los ciudadanos y a los profesionales sanitarios el acceso a la información clínica relevante para la atención sanitaria de un paciente desde cualquier lugar del SNS. La implantación de la Historia Clínica genera una enorme cantidad de datos no estructurados, casi 40 millones de Historias Clínicas Electrónicas que se registraron en España en el 2013 y más de un 1 millón de GB que contienen las imágenes generadas en exploraciones médicas.”

(Menasalvas, Ernestina; Gonzalo, Consuelo; Rodríguez-González, s. f.)

Por último, se hará referencia al, en mi opinión, más importante de los usos que se puede hacer del análisis de datos en el campo de la medicina, que es el estudio de factores de riesgo y prevención y detección precoz de enfermedades y patologías.

Según (Parra Calderón, 2016), existen técnicas que ya se están aplicando con mayor o menor grado de éxito para el hallazgo de factores de riesgo. Al igual que sucede con el seguimiento y la vigilancia de enfermedades, también se ha demostrado que los métodos de Big Data y análisis de datos proporcionan una muy valiosa información acerca de los efectos secundarios adversos que presentan algunos medicamentos.

2.3.2. Agricultura

Es posible que si nos preguntan qué utilidades podría tener la introducción de las herramientas de análisis de datos en el mundo de la agricultura no se nos ocurran demasiadas, pero lo cierto es que existen muchas. Además, este campo cobra una especial relevancia en el caso de España, un país donde la agricultura tiene un peso importante en comparación con otros países de nuestro entorno; y dentro de España, aún tiene un peso mayor en regiones como Castilla y León.

(Castro Franco & Domenech, 2014) nos recuerdan que en el mundo de la agricultura siempre se ha considerado que generar información a partir de los datos puede ayudar a tomar mejores decisiones. Siguiendo estas ideas, constantemente se está persiguiendo innovar en nuevas fuentes de información: hardware, software, maquinaria e instrumentos agrícolas, prácticas novedosas de manejo de cultivos, análisis y estudio de suelos y terrenos, imágenes vía satélite, modelos y sistemas digitales de elevación, cartografía digital de terrenos y suelos, estaciones meteorológicas, análisis financieros y económicos y estudios comerciales y de clientes, entre otras.

En la siguiente ilustración se puede ver cómo la agricultura de precisión, basada en el uso de imágenes aéreas y del empleo de sensores de diversas características, permite registrar una serie de información fundamental para optimizar la gestión del suelo.

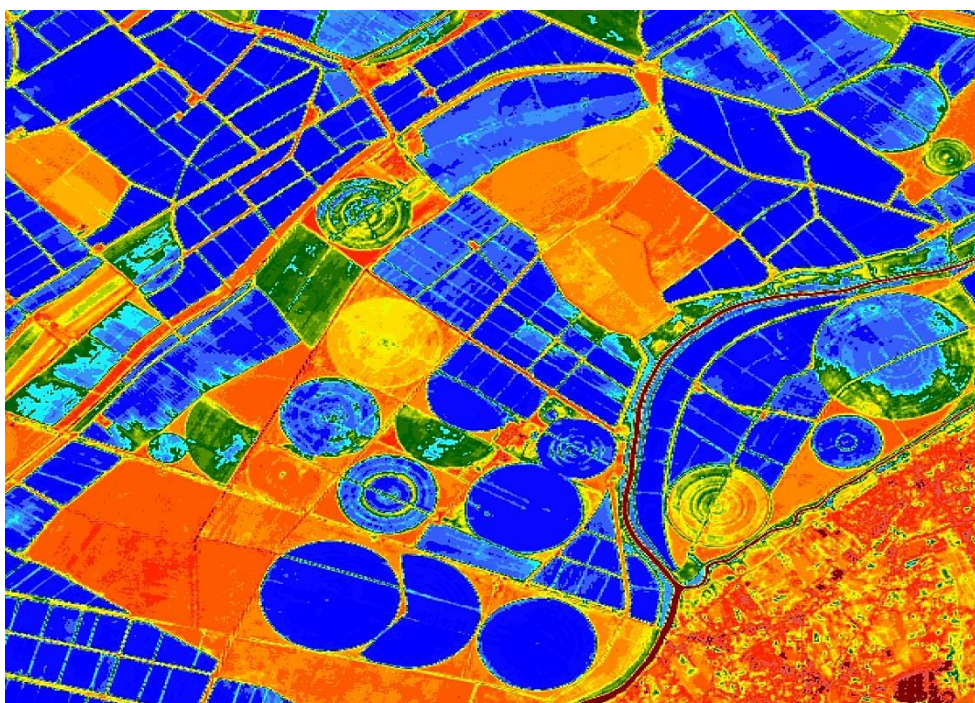


Ilustración 3. Agricultura de Precisión con Imágenes de Satélite, Sensorización y Big Data. Fuente: (Feval Institución Ferial de Extremadura, 2018)

Optimizar el uso del suelo o conocer el estado de las plantaciones son algunas de las posibilidades que el Big Data pone al alcance de la agricultura. Lógicamente, esto traería ventajas no solo para los agricultores y productores, sino también para los consumidores.

“Así como la biotecnología permitió mejorar las semillas y la maquinaria agrícola permitió aumentar la eficiencia en el uso de insumos, las soluciones “Agro Big Data” nos ayudarán a ver el sentido de un nuevo mundo de una forma que apenas intuimos. [...] Se buscará generar herramientas de análisis de datos masivos que permitan generar innovadores conocimientos agronómicos.



El aprovechamiento de los datos masivos abrirá la puerta a nuevas formas de comprender la agricultura. Disponer de datos masivos nos ofrecerá más libertad para explorar, estudiar y examinar más de cerca nuestros sistemas agro-productivos. Además, nos permitirá remodelar el modo en que trabajamos y pensamos la agricultura.”

(Castro Franco & Domenech, 2014)

2.3.3. Comunicación y periodismo

“Es evidente que todo el periodismo tiene como base la información, pero al usar la palabra "datos" o "data", implícitamente nos referimos a un tipo de información particular que puede ser procesada por sistemas informáticos y potentes herramientas de computación (software). De ahí que el periodismo de datos consiste en usar herramientas estadísticas y de visualización para contar mejor las viejas historias y descubrir nuevas historias que contar. Es, en opinión de muchos, la nueva veta del periodismo de investigación. Pero, el desarrollo y puesta en práctica de este tipo de periodismo requiere del conocimiento de otras disciplinas cuyos aprendizajes deben darse transversalmente en los planes de estudios de periodismo.”

(Flores Vivar, 2018)

El mundo de las redes sociales merece una mención especial en este apartado, pues ha revolucionado por completo la forma de comunicar y transmitir información, tanto por parte de los medios como por parte de los usuarios. Este hecho, unido a las constantemente actualizadas ediciones digitales de los periódicos, ha provocado el dramático descenso de las ventas de periódicos en papel.

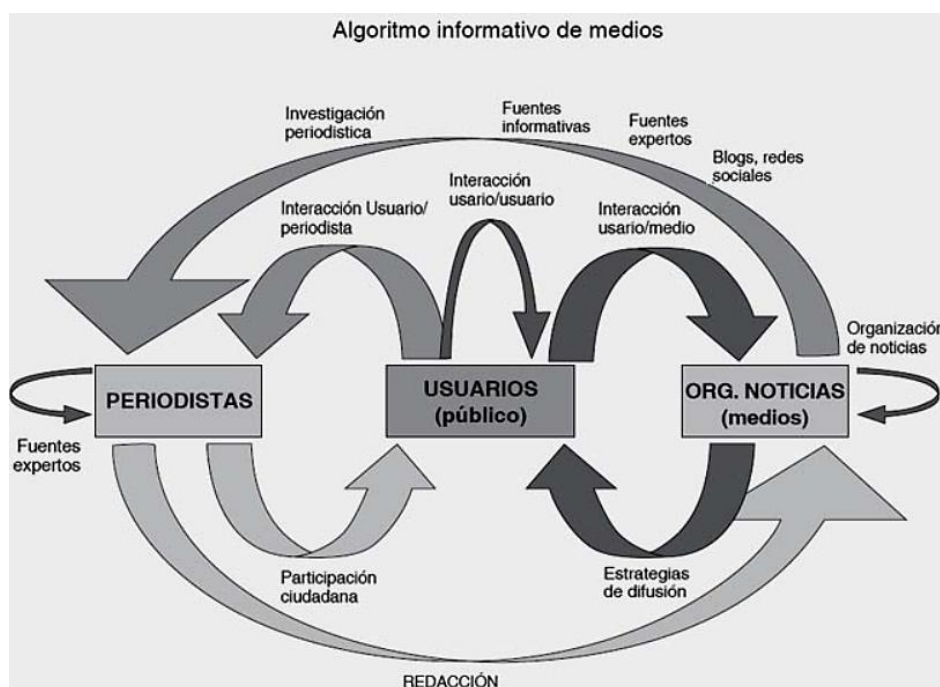


Ilustración 4. Modelo de flujograma algorítmico aplicado a los medios. Fuente: (Flores Vivar, 2018)



Quiero acabar este punto indicando que, los periodistas del futuro deberán tener conocimientos básicos de informática y tecnología, además de ser expertos en investigación en la red.

2.3.4. Desarrollo local y sector público. Países en vías de desarrollo

Es probable que muchos de los avances científicos y tecnológicos que han surgido a lo largo de la historia lo hayan hecho para un marco concreto (el sector militar, el mundo empresarial, el sector aeroespacial...), pero cuando queda demostrado su potencial, se busca su introducción y aplicación en otros sectores, para que todos puedan beneficiarse de estos avances. El caso de las herramientas de análisis de datos y de Big Data no es diferente.

Estamos viendo cómo entornos tan dispares como la medicina y la agricultura o la administración pública y el deporte se aprovechan de los datos y de la información que generan para sacar más provecho a sus áreas y mejorar sus procesos. Por lo tanto, resulta sencillo imaginar que estas herramientas también podrán ser de utilidad para impulsar el desarrollo local de pueblos, comarcas, países subdesarrollados y en vías de desarrollo, etc.

Sin embargo, antes de continuar, quiero resaltar que este desarrollo no será fácil ni estará exento de problemas, tal y como se puede leer a continuación:

“Cuestiones tales como la privacidad y los aspectos técnicos/tecnológicos se encuentran agravados en los países en desarrollo por otros factores, como el atraso tecnológico, la falta de infraestructura y la carencia de personas capacitadas. Este nuevo paradigma se desarrollaría a través de un proceso de difusión lento y desigual, comprometido por la falta de infraestructura, de capital humano y la ausencia de disponibilidad de recursos económicos e institucionales en los países en desarrollo. Esto abre una nueva brecha digital basada en el análisis de datos para la toma de decisiones inteligentes. En consecuencia, la mejora en el desarrollo económico y social no se da automáticamente, sino que requiere elaborar estrategias y políticas públicas específicas.”

(Malvicino & Yoguel, 2016)

Los campos referentes al desarrollo local y al sector público que se pueden abordar son innumerables. Algunos de ellos son: el transporte (optimización del uso de las vías, gestión de los aparcamientos públicos, información de los paneles electrónicos, informatización de los trámites de tráfico, etc.), la gestión de los recursos naturales y la energía (control de la generación, obtención y uso, estudios meteorológicos y seguimiento del clima, importación y exportación de recursos...) y el impulso de determinados sectores en el medio rural.

Sin embargo, es fundamental resaltar que todos estos beneficios no llegarán por sí solos, sino que es necesario tener voluntad y trabajo por parte de los gobernantes y responsables públicos.



(Malvicino & Yoguel, 2016) también relatan en su artículo que conseguir materializar los latentes beneficios para los países que se encuentran en vías de desarrollo requiere la confección e implementación de políticas activas y específicas que valoren la importancia de un adecuado manejo de la privacidad en lo que al tratamiento de los datos personales se refiere, el desarrollo de mecanismos que capaciten y fomenten la creación de valor y la difusión y transmisión de la información, así como la proliferación del conocimiento, con el objetivo final de contribuir a la reducción y, en última instancia, desaparición de las desigualdades socioeconómicas de estos países.

2.3.5. Turismo

Aunque incluya al turismo en el último punto de este apartado del trabajo, este es un sector que, en el caso de España, donde tanta importancia tiene, merece un importante papel dentro de las prioridades de los organismos públicos. La inversión en herramientas de Big Data y análisis de datos en este sector puede suponer la viabilidad de miles de puestos de trabajo en los próximos años y el enriquecimiento de muchas regiones y comarcas de España. En este sentido, el turismo deportivo cobra cada vez más importancia.

“El incremento del tiempo libre, y el estilo de vida cada vez más sedentario, han conseguido que, más allá de la dimensión competitiva, el deporte como práctica diaria y vacacional se haya convertido en una actividad muy demandada, bien como forma de turismo activo y de aventura o bajo las denominaciones de turismo deportivo, senderismo o turismo ecológico. [...] Los autores diferencian entre dos conceptos para definir el turismo deportivo. Por un lado, se refieren al deporte turístico asociándolo con las prácticas deportivas que constituyen espectáculos de atracción para turistas y, por otro lado, se denomina turismo deportivo al practicado por los propios turistas.”

(González Gómez & Rubio Gil, 2019)

Sincronizar las reservas de pistas deportivas, optimizar el uso de las instalaciones, analizar los parámetros (edad, sexo, lugar de procedencia...) de los turistas para mejorar el servicio y ampliar la oferta son algunas de las aplicaciones que el Big Data puede ofrecer en el sector del turismo deportivo.

Lógicamente, esto puede ser extendido a todos los tipos de turismo. Las cadenas hoteleras, los bares y restaurantes y todos los negocios ligados al turismo (que en España son muchos y muy variados) pueden mejorar su actividad e incrementar sus beneficios con un correcto análisis de datos. Cada turista aporta una valiosa información (días de estancia, nacional o extranjero, gasto por persona, lugares que frecuenta, actividades que realiza, grado de satisfacción, número de acompañantes, etc.) que bien interpretada puede mostrar al empresario cómo funciona su negocio: qué aspectos debe mejorar y cuáles son sus puntos fuertes, en qué épocas del año puede subir los precios y cuándo debe hacer ofertas y promociones o qué idiomas deben dominar los empleados son algunas de las



cuestiones que se deben responder para mejorar el negocio y a las cuales pueden dar respuesta las herramientas de Big Data.

2.4. BIG DATA Y ANÁLISIS DE DATOS EN EL DEPORTE

Ya se ha hablado de la historia del Big Data y del análisis de datos en el mundo del deporte (especialmente del fútbol, que es el que marca el eje de este Trabajo de Fin de Grado). Sin embargo, no se ha profundizado lo suficiente en la utilidad y las aplicaciones que estas tecnologías pueden ofrecer en esta área.

“En los últimos 20 años, los datos se han convertido en parte fundamental de las actividades deportivas y el ‘big data’ ya está avanzando con fuerza para sacudir el mundo de los atletas.

Por ejemplo, los equipos de la NBA utilizan actualmente los datos para preparar la estrategia de cada partido, mientras que la NFL cuenta con una plataforma que apoya con sus aplicaciones a todos sus equipos para tomar las mejores decisiones con base en los datos: desde el estado de la superficie del césped y las condiciones climatológicas hasta los datos de la etapa universitaria de cada jugador... todo está registrado y todo puede servir para sacar conclusiones diversas, incluso ayuda a prevenir lesiones de los jugadores.”

(BBVA, 2018)

Prevenir lesiones de los jugadores es, en mi opinión, uno de los puntos más importantes del Big Data en el deporte. Es un secreto a voces que el deporte de élite no es lo mejor para la salud. En algunas disciplinas deportivas, se lleva el cuerpo tan al límite que los deportistas acaban con lesiones crónicas que los lastran durante el resto de su vida. Personalidades del mundo del tenis y del atletismo, entre otros, ya se han quejado de este hecho. El Big Data abre la puerta a llevar a cabo acciones como monitorizar las constantes vitales de los deportistas, controlar su estado del sueño, analizar el grado de sobrecarga de sus músculos y articulaciones u optimizar el tiempo de recuperación de una determinada lesión. Los entrenadores y preparadores físicos, al disponer de esta información, tienen más facilidades para decidir cuándo rotar a un jugador o determinar cuantos minutos puede jugar sin aumentar significativamente el riesgo de lesión.

Por otro lado, se ha hablado de la planificación de los partidos. A la hora de gestionar un equipo deportivo, son cientos los factores que hay que tener en cuenta, pues no es tan fácil como poner a los mejores. Analizando las estadísticas del equipo adversario (disparos por partido, posesión media, tarjetas amarillas y rojas que recibe por partido, goles por encuentro, goles encajados, distribución de sus jugadores en el terreno de juego, etc.) se puede decidir cuál es la mejor estrategia para afrontar el partido. Dependiendo de la forma en la que juega el rival, puede ser más conveniente poner de titular a uno u otro jugador, sacar más o menos delanteros o adelantar o retrasar la defensa, entre otros muchos aspectos.



Todo puede influir y controlar hasta el más mínimo detalle puede suponer una ventaja importante. Por ejemplo, si el partido se va a jugar con una temperatura muy elevada, es posible que un jugador africano o mediterráneo se adapte mejor que un jugador nórdico. Con total seguridad se puede afirmar que, si todas estas consideraciones las tuviera que valorar una persona (o un grupo de personas), muchas cosas se pasarían por alto y no se tendrían en cuenta. Sin embargo, cada día existen más aplicaciones y herramientas que pueden analizar todos estos parámetros y ayudar al equipo técnico a tomar las mejores decisiones estratégicas.

Resalto la palabra ayudar, porque debemos tener presente que las herramientas de Big Data y análisis de datos no van a ganar el partido por nosotros y no van a tomar las decisiones por nosotros. Los responsables últimos seguirán siendo los jugadores, entrenadores, preparadores deportivos y los directivos del equipo. Además, es necesario mencionar que la posibilidad de equivocarse se podrá reducir mucho, pero jamás se podrá eliminar. El deporte tiene un grado de azar, e incluso de suerte, muy importante, lo cual posiblemente hace que sea tan atractivo y que guste tanto. Es parte de su encanto.

“Cada día se crean nuevas métricas para medir el rendimiento de los deportistas y tratar de potenciarlo. [...] En Perú, Ricardo Gareca, técnico de la selección nacional que jugará el Mundial de Rusia 2018, recurrió a la tecnología para analizar en detalle la capacidad de sus jugadores, factor que lo ayudó a tomar la decisión de renovar el equipo nacional y apostar por los jóvenes que clasificaron a la Copa del Mundo tras 36 años de sequía.

En la actualidad, se genera una cantidad de datos enorme, que pueden ser medidos y que permiten introducir cambios en la preparación de los deportistas a partir de la información que se recopila. Todos los datos pueden ser analizados y desmenuzados, desde la distancia en los tiros libres, la cantidad de toques de balón, los pases buenos y malos, la distancia recorrida, la velocidad de un atacante y un largo etcétera. Se sabe que en un entrenamiento de una hora se capturan y analizan 77,7 millones de puntos de datos.”

(BBVA, 2018)

Con relación a lo comentado unos párrafos atrás, en el artículo de (Big Data International Campus, 2016) se destaca que el ojo humano tan solo es capaz de retener en torno al 30% de lo que sucede durante un partido. Por consiguiente, el Big Data permite ayudar a extraer e interpretar conclusiones relacionadas con el juego, a raíz de toda aquella información que es recabada durante un encuentro. Es por eso por lo que el Big Data aporta una gran valía, de la que se pueden aprovechar todos los clubes del mundo, los cuales han comenzado a darse cuenta de ello y están empezando a utilizar estas tecnologías, cuya disrupción en la industria del deporte será exponencial en los próximos años.



Además, tal y como se puede leer a continuación, las áreas de aplicación son muchas y muy variadas:

Como conclusión, se puede decir que “los estudios actuales deben abordar un estudio multivariante que incorpore no sólo indicadores físicos, sino también técnicos, tácticos y psicológicos, pudiendo de este modo abordar una aproximación más completa a la complejidad del rendimiento en competición y entrenamientos. En primer lugar, hay que destacar la creciente importancia de este tópico de investigación [...] por el avance en métodos estadísticos aplicados y por otro lado por el incremento del interés de los clubes, deportistas, medios de comunicación, entrenadores, aficionados, etc. por conocer cómo rinden los jugadores y la explicación que se puede dar a su rendimiento tanto en la competición como en los entrenamientos.”

(Gómez-Ruano, 2017)

2.4.1. El futuro del Big Data en el fútbol

Para terminar este apartado, se expondrán algunas de las aplicaciones y usos del Big Data que ya se están empezando a dar en el mundo del fútbol y que, sin duda, marcarán el futuro de esta tecnología en el deporte. Algunas de ellas fueron expuestas en la conferencia de (González Rodríguez, 2020).

El primer campo de aplicación es el de la ropa. En el deporte de élite, como sabemos, son los pequeños detalles los que marcan la diferencia. Ha habido nadadores que han ganado varias medallas olímpicas por llevar trajes con materiales más novedosos o con texturas especiales. Casos similares se han dado en las equipaciones de los ciclistas y de los atletas. Por lo tanto, analizando la piel de los futbolistas, las partes que más se estiran, las que más sudor producen, las más fácilmente irritables, etc. se pueden diseñar prendas más cómodas y que mejor se adapten al cuerpo; con las ventajas que ello conlleva: mayor facilidad para saltar, tirarse, correr o estirarse.

Por supuesto, merece una mención especial el caso de las botas de fútbol. Estudiando la pisada de cada futbolista, se pueden diseñar y fabricar plantillas personalizadas para cada pie. Esto podría reducir el número de lesiones relacionadas con torceduras de tobillo y mala pisada, los problemas musculares en las piernas y los problemas articulares en tobillos y rodillas.

Otro campo de análisis, que guarda especial relación con la ingeniería industrial, es el de los balones. Se pueden estudiar multitud de materiales y tejidos, para ver cuales tienen mejor comportamiento, en términos de potencia, velocidad y dirección, al ser pateados. Esto mejoraría la precisión en los golpitos, lo que facilitaría las labores de preparación técnica de los futbolistas y añadiría un ingrediente más al espectáculo deportivo. Además, se puede analizar la composición de los balones, con el objetivo de encontrar materiales que ofrezcan las mismas prestaciones con un menor precio, bien por el coste de la materia prima o bien por el coste de



tratamiento del material; o que ofrezcan un mejor rendimiento y/o una mayor durabilidad.

Otra área que presenta un amplio margen de mejora es el de la extracción u obtención de los datos. Mejorando las técnicas que se emplean actualmente para la obtención de los datos y refinando los procesos de tratamiento de estos, se podrá sacar aún mucho más provecho del Big Data. Se podría llegar al caso en el que las cámaras que se encuentran colocadas en los estadios de fútbol para registrar las coordenadas de los futbolistas y los eventos que suceden durante el partido o los sensores que llevan los futbolistas (normalmente en las botas), dieran la información en tiempo prácticamente real, con las ventajas que conllevaría en materia de prevención de lesiones u optimización del juego, entre otras. Aunque hay que decir que esto reduciría parte de la incertidumbre del juego, que es uno de los principales factores que hacen que el fútbol tenga tanto éxito. Por lo tanto, esto no estaría exento de debates y habría que analizar los pros y contras de su aplicación, y cómo ésta se llevaría a cabo. En el próximo apartado se hablará de los datos sin procesar o datos primarios.

Por último, se está valorando la posibilidad de crear bases de datos abiertos para todos los clubes de una liga, para que compartan información que pueda mejorar la competición e impulse el crecimiento y desarrollo de todos los equipos. Por un lado, esto busca mejorar una liga por encima de las de otros países, aunque resulta complicado pensar que, a nivel interno, los equipos estén dispuestos a compartir información con sus adversarios directos. También se habla sobre los datos abiertos en el próximo apartado.

2.5. CONCEPTOS ASOCIADOS AL BIG DATA Y AL ANÁLISIS DE DATOS

Hasta ahora, hemos visto las definiciones, características, objetivos y ventajas del análisis de datos y el Big Data, así como las diferencias y semejanzas entre ambos.

Además, se ha planteado una breve exposición de la evolución histórica de ambos conceptos, así como algunas líneas de desarrollo futuro. Dentro del marco en el que se contextualiza este trabajo, se ha visto la historia de estas tecnologías en el mundo del deporte, donde se ha explicado como su vinculación cobra más y más importancia año tras año. También se han expuesto cuales son los problemas que generan y qué debates hay al respecto.

Por último, he pretendido mostrar el enorme potencial y la enorme utilidad de estas herramientas comentando algunos ejemplos de aplicación, haciendo hincapié en el mundo del deporte. Con ello, debería haber quedado probado que una correcta utilización de las aplicaciones y herramientas de análisis de datos y Big Data pueden ayudar a mejorar y potenciar prácticamente cualquier sector de la sociedad.

Sin embargo, todo esto no es suficiente para comprender como funciona el mundo de los datos y de la información. Por ello, en este apartado plantearé unas



breves exposiciones sobre conceptos y términos asociados al campo del análisis de datos y del Big Data que, desde el punto de vista de la Ingeniería de Organización, ayuden a entender mejor su funcionamiento.

2.5.1. Small Data

Hasta ahora hemos hablado de la gran cantidad de datos que se pueden manejar con el Big Data, así como de la velocidad con la que podemos procesarlos, pero ¿qué sucede si no podemos (por ejemplo, por falta de recursos económicos), no necesitamos (un negocio pequeño) o no sabemos manejar herramientas de tales características? Para ello está el Small Data.

“El Small Data se puede definir como el proceso a través del cual utilizamos y analizamos datos de menor tamaño y mayor simplicidad que pueden afectar a las peculiaridades de una organización. Se trata de datos “simples” que forman parte de la rutina diaria de una empresa y que permiten un conocimiento más profundo y preciso sobre su entorno.

Este análisis pormenorizado de la información ofrece una vertiente cualitativa al análisis de datos que permite una mayor detección de oportunidades para incrementar la eficiencia de los procesos de negocio y estrechar la relación con los clientes, a través del conocimiento preciso de lo que buscan, necesitan y prefieren.”

(Chicano Tejada, 2017)

En otras palabras, podríamos decir que el Big Data pretende prever patrones a gran escala, mientras que el Small Data “ataca” a áreas clave más concretas, con el objetivo de tomar decisiones de un modo eficiente. El Small Data es una herramienta clave para detectar oportunidades de mejora. Es obligatorio mencionar que el coste asociado será menor que el del Big Data. Sin embargo, para optimizar nuestro negocio no deberíamos elegir entre herramientas de uno u otro tipo, sino combinarlas. Small Data y Big Data son tecnologías complementarias.

Tal y como destaca (Chicano Tejada, 2017), es imprescindible operar y examinar grandes volúmenes de datos e información si queremos decretar posibles patrones de comportamiento. Sin embargo, no debemos ignorar que es en los pequeños detalles donde generalmente radica la principal diferencia entre un cliente que momentáneamente queda satisfecho y un cliente fidelizado.

Para terminar, quiero destacar dos puntos acerca del Small Data. En primer lugar, éste está enfocado en el consumidor de una forma más directa. No analiza a todos los consumidores como un gran conjunto, sino que trata de personalizar las necesidades y gustos de cada uno, lo cual es muy útil si, por ejemplo, queremos lanzar campañas más personalizadas.

Y, en segundo lugar, se debe destacar que “aumenta el retorno de la inversión. La inversión en Small Data comparado con el retorno que te ofrece al poder



plantear campañas más eficaces en menos tiempo es mínima; por este motivo el Small Data ayuda a mejorar el ROI.” (ITELLIGENT, 2019)

2.5.2. Datos Abiertos (*Open Data*)

“Otro concepto empleado en la actualidad es el de Open Data; que aboga por la libertad de una serie de datos para que puedan ser empleados por cualquier ciudadano. Entre estos datos destacarían las encuestas de la EPA, el paro o el IPC, entre otros.”

(Antón Carranza, 2017)

En los últimos años, son cada vez más las voces que opinan que hay datos que deben ser públicos y estar accesibles para todo el mundo. Desde el punto de vista de la ciencia, se cree que esto alienta la investigación y el desarrollo, maximizando el beneficio para la sociedad. En el caso de la medicina sucede lo mismo, si los avances que se producen son conocidos prácticamente en tiempo real en todos los rincones del mundo, serán muchas más las personas que se beneficien de ellos y, a mayores, las investigaciones se desarrollarían de una forma mucho más rápida. En el ámbito de la administración pública y gubernamental, el concepto de transparencia va en esta línea. E incluso hay quienes opinan que las empresas privadas se beneficiarían si compartieran los datos las unas con las otras.

Todos estos argumentos se consolidan cuando la información proviene de una investigación llevada a cabo con dinero público o realizada directamente por una institución pública. A continuación, se ofrece una definición lo más objetiva posible de lo que son los datos abiertos:

“El concepto datos abiertos (open data, en inglés) es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control.”

(Wikipedia, 2020b)

El movimiento de Datos Abiertos va en línea con otros movimientos de ética similar, como el de software libre, el de código abierto y el de acceso abierto. Todas estas corrientes presentan un enorme atractivo, también desde el punto de vista de la ingeniería en organización industrial. Sin embargo, no se versará más sobre ello, puesto que no es el objeto principal de estudio de este trabajo y se aleja de los objetivos de conocimiento que se pretenden alcanzar.

Para terminar, se hablará sobre las características fundamentales de los datos abiertos:

“- Disponibilidad y acceso. Los datos deben encontrarse disponibles, preferiblemente mediante la descarga a través de internet, y permitir su modificación. Su reproducción debe ser a un coste razonable.



- *Reutilización y redistribución en condiciones óptimas pudiendo combinarse con otros conjuntos de datos.*

- *Participación universal: todos deben ser capaces de utilizarlos, reutilizarlos y redistribuirlos. No deben existir restricciones para su comercialización.”*

(Ferrer-Sapena & Sánchez-Pérez, 2013)

2.5.3. Datos sin Procesar (Raw Data)

Los datos sin procesar, también conocidos como datos primarios, son aquellos que han sido recopilados de una fuente y se encuentran tal cual se recogieron, sin haber sufrido ninguna manipulación ni haber sido sometidos a algún tipo de procesamiento.

“Los datos sin procesar generalmente se refieren a tablas de datos donde cada fila contiene una observación y cada columna representa una variable que describe alguna propiedad de cada observación. Los datos en este formato a veces se denominan datos ordenados, datos planos, datos primarios, datos atómicos y datos de registro de unidades.”

(Bock, 2018)

En relación con el análisis de datos en el mundo del fútbol, (González Rodríguez, 2020) explicaba que los datos brutos recogen toda la información que sucede durante los 90 minutos de juego. Se trata de un fichero XML (en el caso del Real Valladolid CF y de otros equipos de la Liga Española de Fútbol, pero podría hallarse en otros formatos) en el que se recogen las posiciones según coordenadas XYZ de cada jugador, el árbitro y los árbitros asistentes y el balón.

(González Rodríguez, 2020) arrojaba, además, algunos datos: Durante un partido de fútbol se crean, aproximadamente, unos 80 millones de líneas de código. Para poder emular esto, es decir, para recopilar la misma cantidad de información, se necesitarían 200.000 ojeadores por partido. Si quisiéramos imprimir dicha información necesitaríamos 20.000 kilómetros de papel, para una temporada con 22 equipos. Terminaba afirmando que el raw data es lo más importante dentro del Big Data en el fútbol.

2.5.4. Tipos de bases de datos: Relacional y no relacional

Conocer cuál es y cómo funciona la base de datos que se utiliza en un estudio o una investigación es fundamental para el éxito de ésta. Los problemas a la hora de interpretar, modificar o ampliar una base de datos son bastante frecuentes y pueden arruinar un buen trabajo. Sin profundizar demasiado, se explicará qué tipos de bases de datos existen, cuáles son sus características, ventajas e inconvenientes y cuáles son sus semejanzas y diferencias.

“La base de datos es el concepto que hace referencia a cómo está organizada una información, que puede ser de estructura relacional o no relacional, e influye determinantemente en el tratamiento de los datos.



El concepto de base de datos relacional es una estructura de datos que permite que la información esté relacionada entre sí, entre distintas 'tablas' o diferentes tipos de contenedores de datos. Un contenedor de datos debe contener una clave primaria que permita conectar de manera unívoca con cada instancia del contenedor. Una base de datos no relacional almacena la información sin una estructura específica ni mecanismo de enlace de los datos de un contenedor a otro definido de base."

(González-Iglesias González, 2015)

Una vez hemos visto la definición de ambos tipos, veremos las ventajas e inconvenientes que presentan cada uno de ellos. Se empezará por las bases de datos relacionales. Sus ventajas son las siguientes:

"- Al ser una tecnología bastante madura, cuenta con una documentación muy extensa y una comunidad bastante activa. Cualquier duda puede ser resuelta con un poco de investigación.

- Los estándares SQL se encuentran bien definidos y son ampliamente aceptados.

- Una gran cantidad de desarrolladores cuentan con amplia experiencia en esta tecnología.

- Toda base de datos relacional debe cumplir con los principios ACID, por lo cual los datos son confiables."

(Cid Robles, 2019)

Los principios ACID hacen referencia a que la base de datos siga las reglas de atomicidad (todos los pasos se ejecutan sobre una sola operación), consistencia (la operación debe acabar en un estado válido), aislamiento (una operación no puede afectar a otras) y durabilidad (una operación debe ser permanente y el último estado conocido de los datos debe ser almacenado).

Con ello, se procede a estudiar algunas de las desventajas que presentan las bases de datos relacionales, según (Cid Robles, 2019):

- Tienen muy poca escalabilidad.
- Para que funcionen, los datos deben estar muy bien estructurados.
- Migrar de un sistema gestor a otro es un proceso muy complejo.

Por su parte, (Cid Robles, 2019) también habla de algunas de las ventajas de las bases de datos no relaciones, que son las que se muestran a continuación:

- Este tipo de BBDD son sumamente escalables, con lo que se pueden editar al mismo tiempo que el código que vamos creando.
- También son bastante tolerantes a errores y fallos.
- Se considera que son bases de datos fáciles de respaldar, ya que es suficiente con crear una copia del archivo.



- Los datos pueden adecuarse a otros tipos de sistemas de una forma sumamente sencilla, debido a que no presentan un esquema específico.

Y los inconvenientes o desventajas que plantea (Cid Robles, 2019) serían:

- Estas bases de datos representan una tecnología relativamente reciente, lo que conlleva que la información y el conocimiento sean más limitados.
- Cada BBDD tiene su manera propia de formatear los datos de los que dispone, a diferencia de lo que sucede con los lenguajes de consulta estructurados, como es el caso de SQL.

Antes de continuar, se ha de destacar que, contrariamente a lo que muchas personas creen, las bases de datos no relacionales no constituyen un reemplazo para las bases de datos relacionales, sino que se pueden entender como un complemento o una alternativa, surgidas para satisfacer los requerimientos o necesidades de las nuevas aplicaciones, que cada vez son más complejas.

Pese a lo ya explicado, aún pueden quedar dudas sobre cuáles son las diferencias más significativas que hay entre los dos tipos. Como respuesta a esta cuestión, podríamos exponer lo siguiente:

Las llamadas bases de datos relacionales se basan en que la información se organiza en pequeñas partes que se integran mediante lo que podríamos denominar identificadores. Dicho postulado, según explica (Alexis Rendón, 2019), las diferencia de las bases de datos no relacionales que, tal y como su propio nombre indica, no presentan un identificador que se encargue de vincular o relacionar dos o más conjuntos de datos. Además, las BBDD relacionales también son más robustas. En otras palabras, disponen de una mayor capacidad de almacenamiento, y son menos vulnerables ante defectos y fallos.

A mayores de lo expuesto, se puede mencionar que las bases de datos no relacionales son más fáciles de desarrollar y tienen un mayor rendimiento. Son muy útiles cuando los datos a almacenar no siguen un esquema o patrón exacto.

En la siguiente tabla se puede ver, de una forma bastante resumida, cuando es más conveniente usar cada tipo de bases de datos; aspecto que, como ya se ha mencionado al principio de este subapartado, es de vital importancia conocer y dominar.



SQL	NoSQL
<p>Cuando el volumen de mis datos no crece o lo hace poco a poco.</p> <p>Cuando las necesidades de proceso se pueden asumir en un sólo servidor.</p> <p>Cuando no tenemos picos de uso del sistema por parte de los usuarios más allá de los previstos.</p>	<p>Cuando el volumen de mis datos crece muy rápidamente en momentos puntuales.</p> <p>Cuando las necesidades de proceso no se pueden prever.</p> <p>Cuando tenemos picos de uso del sistema por parte de los usuarios en múltiples ocasiones.</p>

Tabla 2. Cuando usar una Base de Datos relacional y no relacional. Fuente: (Alexis Rendón, 2019)

Como conclusión a este subapartado, se puede decir que, a día de hoy, las bases de datos relacionales continúan siendo las predominantes. Sin embargo, las bases de datos no relacionales están cobrando cada vez más importancia. No las hay mejores ni peores, sino que el tipo a elegir deberá venir determinado por las características de nuestro negocio y las necesidades específicas que debemos cubrir. Por último, mostraré un esquema típico de una base de datos relacional, para que se pueda entender el concepto de una forma más visual.

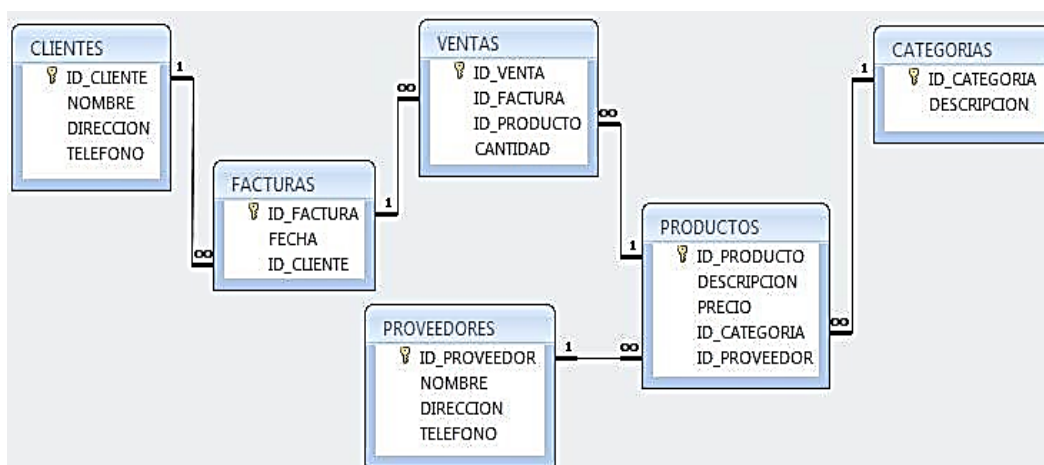


Ilustración 5. Esquema de una Base de Datos Relacional. Fuente: (TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN II (TIC'S II), s. f.)

2.5.5. Tipos de recogida de datos: Clasificación IBM

Se ha hablado sobre los tipos de bases de datos, que son los espacios donde se almacena la información. Como se ha comentado, es muy importante elegir bien la base de datos a emplear y conocer a la perfección su funcionamiento. Sin embargo, antes de ese paso, deberíamos preguntarnos: ¿de dónde proceden los datos que queremos almacenar?

Poder responder a esta cuestión es fundamental para la validez del estudio o la investigación, pues debemos garantizar que la fuente (o las fuentes) sean fiables,



que la información este contrastada y actualizada y que ésta sea coherente y uniforme. Tal y como detalla (Costalago Serrano, 2019), según la empresa IBM existen cinco tipos diferentes de recogida de datos. Son los siguientes:

Webs y redes sociales: En este grupo incluiríamos toda la información obtenida de las páginas web y de las redes sociales. Como todo el mundo supondrá, ésta es la mayor fuente de datos que existe en la actualidad. Los usuarios cuelgan en las redes sociales sus opiniones sobre productos, servicios, lugares, personas, política, deporte, etc. además de informar a diario de que series y películas ven, que libros leen, que cosas han comprado o a donde van a ir en sus próximas vacaciones. Por lo tanto, las redes sociales suponen una inagotable fuente de información que, además, cuenta con la ventaja de mostrarla en tiempo real. Lo mismo sucede con las páginas webs, donde las empresas pueden ver qué interacciones realizan los clientes, qué productos son los más demandados, qué opiniones se dan sobre cada uno...

Información máquina a máquina: Aquí se incluyen aquellos datos que han sido generados, recopilados o capturados por una máquina y que son almacenados y/o transmitidos a través de redes inalámbricas o híbridas. Este tipo de recopilación de datos se ha potenciado mucho en los últimos años, debido a la aparición de tecnologías que permiten conectar unos dispositivos con otros y que éstos se comuniquen e interaccionen entre sí.

Grandes transacciones de datos: “En este apartado se recogería toda la información inherente a las transacciones masivas de información, como podrían ser los centros de atención telefónica, de banca, finanzas o de atención al cliente entre otros tipos” (Costalago Serrano, 2019). Este grupo presenta áreas comunes con el de webs y redes sociales; y su principal diferencia es que los datos obtenidos en las redes sociales tienen un nulo grado de estructuración, mientras que los obtenidos en los servicios telefónicos de atención al cliente o formularios de banca, por ejemplo, sí que presentan ciertos esquemas o patrones.

Biometría: “Aquí se recoge toda la información biométrica, la cual es sumamente importante para organizaciones como los gobiernos, cuerpos de policía, seguridad privada, servicios de inteligencia, etc.” (Costalago Serrano, 2019). El incremento de dispositivos que se desbloquean mediante la verificación de la identidad del usuario a través de la huella dactilar, patrones faciales, la retina o el iris es, posiblemente, el más claro ejemplo de este grupo de recogida de datos.

En el mundo del deporte puede adquirir una gran importancia para, por ejemplo, analizar muestras de sangre de un deportista para validar una prueba antidopaje.

Información generada por los seres humanos: Este último tipo engloba todos los datos digitales originados por las personas, de una forma general.



2.5.6. Árboles de decisión

Para comprender adecuadamente qué son los árboles de decisión, en primer lugar, se presentarán unas definiciones de éstos. A continuación, se comentarán algunas de sus características y ventajas y se mencionarán los elementos que forman un árbol de decisión. Por último, se mostrará un ejemplo visual. Como definiciones podríamos exponer las siguientes:

Los árboles de decisión son “modelos de decisión basados en grafos de estructura tipo árbol, en los que en cada rama se evalúa el resultado de segmentar el conjunto a estudio por una dicotomía de un caso por una variable independiente.”

(González-Iglesias González, 2015)

Otra posible definición, más intuitiva, es la siguiente:

“Los árboles de decisión son un método de aprendizaje supervisado no paramétrico, utilizado para la clasificación y regresión. El objetivo es crear un modelo que prediga el valor de una variable objetivo mediante el aprendizaje de reglas de decisión simples inferidas a partir de las características de los datos. Los árboles de decisión funcionan al evaluar una expresión que contiene una característica en cada nodo y seleccionar una rama al siguiente nodo en función de la respuesta.”

(Leguizamón Rojas, 2018)

Por último, una tercera forma de definir este concepto es la siguiente:

“Un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más “acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Estos árboles permiten examinar los resultados y determinar visualmente cómo fluye el modelo. Los resultados visuales ayudan a buscar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales. Los árboles de decisión son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas.”

(Berlanga Silvente et al., 2013)

Con las definiciones propuestas, podemos hacernos idea de lo que son los árboles de decisión y de la utilidad que ofrecen. En lo referente a sus características y ventajas, se pueden destacar los siguientes aspectos:

En primer lugar, se debe decir que las principales ventajas que ofrecen los árboles de decisión son, tal y como comenta (Leguizamón Rojas, 2018), que son sencillos de entender e interpretar, que son aptos para realizar de forma implícita la



selección de características, y que pueden operar tanto con datos numéricos como con datos categóricos. A mayores, (Leguizamón Rojas, 2018) dice que requieren un esfuerzo relativamente escaso por parte de los usuarios, de cara a la preparación de los datos. Los nexos no lineales que hay entre los parámetros no afectan al rendimiento operativo del árbol.

Si ahora se habla de caracterizar a los árboles de decisión, (Berlanga Silvente et al., 2013) relatan que son una técnica o procedimiento de minería de datos, cuya función es preparar, explorar y sondear los datos disponibles para extraer la información que se encuentra oculta en ellos. Con ello, se aborda la solución a los problemas de predicción, segmentación y clasificación, entre otros.

“La construcción recursiva del árbol se detiene cuando se alcanza la profundidad máxima, cuando no se encuentra un atributo que aumente la ganancia de información; o cuando ningún atributo obtenga nodos hijos que sean alcanzados por un número mínimo de instancias del conjunto de entrenamiento.”

(Leguizamón Rojas, 2018)

Otras ventajas de esta herramienta son las que se citan a continuación:

*- Facilita la interpretación de la decisión adoptada.
- Facilita la comprensión del conocimiento utilizado en la toma de decisiones.
- Explica el comportamiento respecto a una determinada decisión.
- Reduce el número de variables independientes.”*

(Berlanga Silvente et al., 2013)

Sin embargo, también se pueden mencionar algunos inconvenientes y desventajas con respecto a otras herramientas, como se explica a continuación, según el trabajo de (Leguizamón Rojas, 2018):

Los árboles de decisión pueden engendrar árboles demasiado complejos que no generalicen de forma correcta los datos. A este hecho se le conoce como sobreajuste. Estos árboles pueden llegar a ser inestables, lo que provocaría que pequeñas variaciones en los datos dieran como resultado la concepción o generación de un árbol totalmente diferente (lo que en términos estadísticos se conoce como varianza). Como solución, (Leguizamón Rojas, 2018) explica que existe la posibilidad de crear árboles sesgados si se dominan algunos métodos y técnicas.

Dando por finalizada la parte de características, ventajas e inconvenientes, es interesante hablar ahora de los elementos que componen un árbol de decisión, tema que presenta un gran interés y que resulta notablemente importante conocer si queremos trabajar con ellos. Dichos elementos siguen la terminología que se presenta y detalla a continuación:



“Nodo de decisión: Nodo que indica que una decisión necesita tomarse en ese punto del proceso. Está representado por un cuadrado.

Nodo de probabilidad: Nodo que indica que en ese punto del proceso ocurre un evento aleatorio. Probabilidades de que ocurran los eventos posibles como resultado de las decisiones. Está representado por un círculo.

Nodo terminal: Nodo en el que todos los casos tienen el mismo valor para la variable dependiente. Es un nodo homogéneo que no requiere ninguna división adicional, ya que es “puro”.

Rama: Nos muestra los distintos caminos que se pueden emprender cuando tomamos una decisión o bien ocurre algún evento aleatorio. Resultados de las posibles interacciones entre las alternativas de decisión y los eventos.”

(Berlanga Silvente et al., 2013)

Para acabar este punto, tal y como se comentó en el primer párrafo, se muestra un sencillo ejemplo de árbol de decisión:

Ejemplo: Árbol de decisión para “¿ Administrar fármaco F ?”

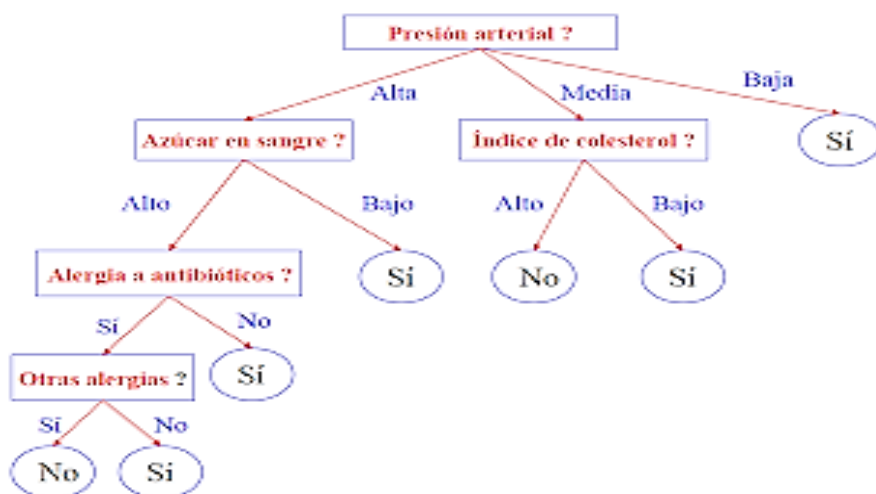


Ilustración 6. Ejemplo de árbol de decisión. Fuente: (Martínez & Ventura, 2016)

2.5.7. ETL

“Extract, Transform and Load («extraer, transformar y cargar», frecuentemente abreviado ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio. Los procesos ETL también se pueden utilizar para la integración con sistemas heredados. Se convirtieron en un concepto popular en los años 1970.”

(Wikipedia, 2020c)

La primera parte del proceso ETL, Extraer, es la que consiste en extraer, valga la redundancia, los datos desde el sistema de origen. Lo más frecuente es que haya



más de un sistema de origen. El principal problema que puede aparecer en esta etapa es que cada sistema de origen presente los datos en un formato diferente.

Con respecto a esto, (González Rodríguez, 2020) comentaba en su conferencia que algunos aspectos del juego del fútbol no se consideraban bajo la misma denominación en España que en Inglaterra, por ejemplo. Por lo tanto, los equipos deben tener especial cuidado al trabajar con las bases de datos y ser conscientes del país de origen de la base.

En esta primera parte se convierten los datos a un único formato, sobre el cual se trabajará en la siguiente etapa.

Esta segunda etapa es la Transformación. En ella, se aplican reglas y/o funciones que transforman los datos. A veces es necesario llevar a cabo una pequeña manipulación de los datos. Algunos ejemplos son: traducir códigos, transponer columnas y filas, obtener valores calculados...

La tercera y última fase es la Carga, consistente en cargar todos los datos tratados en el paso anterior en el sistema de destino.

2.5.8. Inteligencia Artificial

El campo de la inteligencia artificial no sólo es altamente interesante y novedoso, sino que es de una extensión impensable. Constituye un muy atractivo campo de investigación en el que aún está todo por descubrir. Este concepto podría dar para hacer un trabajo de estas características sólo hablando de él, e incluso para escribir tesis y libros.

Puesto que no es el principal objeto de estudio de este TFG, tan sólo se expondrán algunas nociones básicas que permitan entender mejor en qué consiste la inteligencia artificial y cuáles son sus utilidades y aplicaciones. Aun así, se debe resaltar fuertemente que el campo de la inteligencia artificial guarda una intensa relación con la ingeniería industrial.

*“El nacimiento de la inteligencia artificial como una disciplina de investigación científica podríamos situarlo en una conferencia sobre la informática teórica que se desarrolló en el Dartmouth College en los Estados Unidos en 1956. [...] En aquella conferencia, A. Newell y H. Simon expusieron su estudio *Logic Theorist*. Se trataba de un programa de ordenador que emulaba las características que presenta un cerebro humano, y es por eso por lo que se le consideró el primer sistema de inteligencia artificial de la historia. Aquel programa que presentaron Newell y Simon era capaz de demostrar la mayor parte de los teoremas sobre lógica matemática que se habían presentado hasta la época.”*

(Costalago Serrano, 2019)

En este contexto, (Bandrés Balet, 2019) expone en su trabajo que actualmente, debido a la continua evolución y al exponencial desarrollo de la



tecnología, se dispone de la cantidad de información necesaria para que los modelos y sistemas informáticos estén capacitados para actuar y comportarse como si fuesen un ser humano. Este complejo proceso se realiza a través de lo que se conoce como algoritmos inteligentes, los cuales permiten llevar a cabo una toma de decisiones a partir de la interpretación y el entendimiento de los datos disponibles y del conocimiento adquirido de las personas.

Retomando de nuevo una visión histórica de la evolución de la inteligencia artificial, es interesante comentar el siguiente hecho, que tuvo lugar en la década de los años 50:

“Alan Turing publicaba en la revista Mind, Computing machinery and intelligence, un artículo en el que se expone un test de inteligencia que serviría para comprobar si una máquina es capaz de presentar un comportamiento inteligente por sí misma. El estudio trataba de establecer una conversación entre un humano y un ordenador. De esta manera, en la medida que si una persona externa a la conversación no es capaz de distinguir quién es la máquina y quién el ordenador, estaríamos ante una inteligencia artificial real. [...] A pesar de tantos avances, todavía no se ha diseñado una máquina que haya conseguido pasar la prueba de Turing con solvencia. [...] La prioridad de esta ciencia ya no es tanto la posibilidad de construir aparatos capaces de comportarse como un humano y tener sus capacidades, sino la posibilidad de crear sistemas capaces de procesar tales volúmenes de información que escapen a las capacidades del propio ser humano”

(Costalago Serrano, 2019)

Seguramente la mayoría de los expertos no se pongan de acuerdo para dar una definición precisa de lo que es la inteligencia artificial. A continuación, se tratará de dar una descripción lo más completa posible:

“La inteligencia artificial es una disciplina académica relacionada con la teoría de la computación, cuyo principal objetivo es el de emular algunas de las facultades humanas en sistemas artificiales. Con inteligencia humana, nos referimos a procesos de percepción sensorial y a sus consiguientes procesos de reconocimiento de patrones. Por lo que las aplicaciones más habituales de la IA desde sus inicios en los años 60 son el tratamiento de datos y la identificación de sistemas.”

(Benítez et al., 2013)

Con todo lo explicado hasta aquí, podemos hacernos una idea de qué es la inteligencia artificial y cómo ha evolucionado desde sus comienzos. Sin embargo, aún queda dar respuesta a la siguiente cuestión: ¿dónde se aplica esta tecnología en la actualidad y en qué ha triunfado?

Según analiza (Monedero Carreras, 2019), en la actualidad se ha extendido considerablemente el uso de la inteligencia artificial como técnica o método de



mejora y desarrollo de los procesos, tanto en las grandes corporaciones como en las pequeñas y medianas empresas. Los programas y sistemas dotados de IA están capacitados para ejecutar diferentes labores y tareas de una forma más rápida y precisa que los seres humanos. De este modo, según comenta (Monedero Carreras, 2019), se logra una disminución del número de errores y se mejora el flujo de la actividad de la organización. La aparición de estos programas y herramientas ha permitido que el ser humano pueda centrarse en aquellas labores que son más críticas para el negocio y, de esta forma, enriquecer la calidad de los servicios prestados al cliente.

2.5.9. Industria 4.0

Al igual que sucedía en el caso anterior, la industria 4.0 hace referencia a un concepto tan amplio que existen libros para tratarlo e incluso se dan asignaturas en algunas carreras universitarias sobre ello. En este trabajo tan solo se verá una definición que permita ver la innegable relación entre la industria 4.0 y el Big Data y el análisis de datos, cuyas herramientas han permitido y permiten progresar al mundo de la informática, las telecomunicaciones y, por supuesto, también a la industria.

Siguiendo la definición propuesta por (Bandrés Balet, 2019), se puede afirmar que, la denominada Industria 4.0, es una nueva revolución que combina técnicas avanzadas de producción y fabricación con tecnologías perspicaces que día a día se están integrando con mayor fuerza tanto en las organizaciones y empresas de todo el mundo como en las propias personas e incluso en los activos. También resalta (Bandrés Balet, 2019) que la revolución de la Industria 4.0 está marcada por el surgimiento y desarrollo de nuevas ciencias tecnológicas como la robótica, la inteligencia artificial, la nanotecnología, las tecnologías cognitivas y, con una especial relevancia, el conocido como Internet de las Cosas (abreviado *IoT*, por sus siglas en inglés).

Otra descripción bastante interesante y completa que nos puede ayudar a comprender mejor el concepto que se está tratando es la que se muestra a continuación:

“El término industria 4.0 se refiere a un nuevo modelo de organización y de control de la cadena de valor a través del ciclo de vida del producto y a lo largo de los sistemas de fabricación apoyado y hecho posible por las tecnologías de la información. El término industria 4.0 se utiliza de manera generalizada en Europa, si bien se acuñó en Alemania. También es habitual referirse a este concepto con términos como “Fábrica Inteligente” o “Internet industrial”. En definitiva, se trata de la aplicación a la industria del modelo “Internet de las cosas” (IoT). Todos estos términos tienen en común el reconocimiento de que los procesos de fabricación se encuentran en un proceso de transformación digital, una “revolución industrial” producida por el



avance de las tecnologías de la información y, particularmente, de la informática y el software.”

(Del Val Román, 2016)

2.5.10. Efecto Moneyball

El Efecto *Moneyball* es uno de los conceptos más interesantes que se estudian en este Trabajo de Fin de Grado pues, aunque se trata de un término que tiene relación directa con el análisis de datos y con el objetivo de trabajar con ellos para extraer información útil que permita tomar mejores decisiones estratégicas, también se trata de un concepto puramente deportivo, con lo que es idóneo para el supuesto práctico que más adelante se tratará.

Para adentrarnos en el Efecto *Moneyball*, se comenzará narrando la siguiente historia de éxito alcanzado gracias al uso de la estadística en el deporte:

“Una de las historias deportivas que mayor repercusión ha tenido en los últimos años, fue la hazaña lograda por el equipo de béisbol de la ciudad de Oakland, los Oakland Athletics en 2002, cuando su manager, Billy Beane, apostó por la aplicación de las estadísticas avanzadas como herramienta de mejora del rendimiento deportivo de su equipo, lo que condujo a ese modesto club al éxito deportivo.

El béisbol es un deporte más propicio al uso de las estadísticas que el fútbol debido a sus particularidades, pero en materia de fichajes ha funcionado de manera similar durante muchos años: de acuerdo con la subjetividad e intuición de ojeadores y entrenadores. No es que este método no haya funcionado en muchos casos, pero se encontró otro más fiable y eficiente, capaz de reducir las distancias presupuestarias entre los equipos. Fue el estadístico Bill James quien comenzó a utilizar la estadística avanzada para medir el rendimiento de los jugadores.”

(Antón Carranza, 2017)

El novedoso enfoque que se introduce en el mundo del béisbol esa temporada hace que este deporte cambie para siempre. Hasta ese momento, el modelo de seguimiento de los jugadores llevado a cabo por los ojeadores se basaba, según explica (Blanco, 2016), en criterios estrictamente de logros y éxitos personales de cada uno de los jugadores (número de carreras conseguidas, número de golpes acertados, cantidad de bases robadas, etc.). Este hecho fomentaba un modelo de mercado de fichajes inflacionista, lo que provocaba que el reclutamiento de jugadores talentosos fuera difícilmente asumible, a nivel económico, para una gran parte de los equipos de la competición.

Como todos los aficionados al fútbol tendrán presente, este problema sigue vigente actualmente y, en el caso de España, todos los equipos ven como año tras año disminuyen las posibilidades de disputar la victoria en el campeonato nacional de liga al Real Madrid y al FC Barcelona.



Con la introducción de la estadística en el beisbol, se empezaron a analizar aspectos de los jugadores que hasta ese momento habían sido infravalorados, con lo que los clubes más modestos pudieron abordar los fichajes de jugadores cuyo valor era más o menos bajo, pero que los datos reflejaban un rendimiento más o menos bueno. Este ejemplo nos conduce directamente a la definición de lo que es el efecto *moneyball*.

“El empleo de algoritmos matemáticos fue el desencadenante del efecto Moneyball, consistente en contratar jugadores infravalorados por el mercado, pero con un rendimiento superior del esperado.

Fue en el año 2001 cuando dos estudiantes de la Universidad de Yale fueron contratados por el mánager de los Oakland Athletics, Billy Beane, para aplicar ese conocimiento matemático y contratar a bajo coste jugadores de gran nivel, pero denostados por aspectos subjetivos como la edad, el físico o la personalidad. El equipo de Oakland, con uno de los presupuestos más bajos de la liga consiguió batir el récord de victorias consecutivas en una temporada (20) y llegar a la final de la Liga americana.”

(Antón Carranza, 2017)

A modo de conclusión del ejemplo expuesto, se puede valorar la reflexión que hacen (Simmons & Berri, 2019) en su artículo. Los autores comentan que una de las tesis centrales en las que se apoyaba originalmente el Efecto *Moneyball* es que los directores técnicos de los equipos, responsables de tomar las decisiones en materia de fichajes, no estaban valorando correctamente la información que proporcionaban los datos e indicadores de rendimiento de los jugadores de béisbol.

La cuestión de acertar plenamente con los fichajes es fundamental especialmente en los equipos más modestos. Si quieren evolucionar y poder competir algún día por ganar el campeonato o, al menos, quedar en posiciones cómodas alejadas de los puestos de descenso, deben ser capaces de encontrar a jugadores baratos que ofrezcan un buen rendimiento y que permitan, en el futuro, ser vendidos por una cantidad muy superior a la que se pagó por su fichaje. Ésta es la única manera de mejorar al mismo tiempo la sección deportiva y la económica. Posiblemente, el ejemplo más claro que hay en el fútbol es el Sevilla CF, gestionado por el director deportivo Ramón Rodríguez Verdejo, más conocido como *Monchi*, que es conocido por su gran capacidad y habilidad para fichar jugadores desconocidos o semidesconocidos, y sacarles un gran potencial deportivo y económico.

2.5.11. Almacenes de datos (Data Warehouses)

“Aunque William Harvey Inmon es considerado como el padre del Data Warehouse, fueron unos investigadores del IBM quienes le dieron este nombre. Fue creado en la década de los 90, y consiste en un conjunto de datos almacenados que pueden ser consultados por las empresas mediante tecnologías como la Minería de Datos. [...] Estos datos sirven de apoyo para



tomar decisiones y se encuentran ordenados, organizados por temas, son temporales y no volátiles.”

(Rivas González, 2019)

Una definición más completa y técnica, que nos permita entender mejor que son los almacenes de datos o Data Marts, podría ser la siguiente:

“Un Data Warehouse es una base de datos corporativa tradicional, la cual destaca por integrar y depurar la información antes de su almacenamiento, siguiendo procesos ETL, facilitando así su análisis desde un gran número de perspectivas a gran velocidad. [...] Los datos son almacenados con una estructura previa definida.”

(Gutiérrez del Campo, 2019)

Dentro del campo de los almacenes de datos es necesario hablar de los conocidos como mercados de datos, que se podrían describir de la siguiente forma:

“Son un subconjunto de un Data Warehouse orientado al análisis, almacenamiento e integración de los datos en un departamento o área de la empresa. Posee la misma complejidad que el Data Warehouse pero se estructura de manera diferente, ya que cada departamento tiene unas necesidades distintas y una forma de gestionar y aplicar la información.”

(Rivas González, 2019)

La diferencia primordial entre los almacenes de datos y los mercados de datos es que, según analiza (Martín Peña, 2016) en su trabajo, los primeros buscan cubrir aquellas necesidades que se pueden considerar comunes a toda la organización y, por contra, a los mercados de datos les es suficiente con agasajar a tan solo una porción de la empresa (un departamento, una sede o un grupo de trabajo, por ejemplo).

2.5.12. Segmentación de Datos

Como ya se ha comentado en varios apartados, lo más probable es que la cantidad de datos recopilados para un estudio o una investigación sea muy grande; por lo tanto, antes de trabajar con ellos deberemos llevar a cabo una segmentación de datos. Para entender la utilidad de esta fase o etapa se propone un ejemplo del ámbito futbolístico.

Tras finalizar la temporada de fútbol, comienza el período de fichajes, en el que los equipos deben comprar y vender jugadores según sus necesidades (deportivas y/o económicas). Para ello, la dirección técnica deberá obtener información sobre todos los jugadores (datos deportivos, informes de ojeadores, estadísticas...), tanto los que juegan en España como los que juegan en cualquier otra liga del mundo. Salta a la vista que esta información es intratable debido a su enorme tamaño. Se tardarían años en analizar todos los parámetros de todos los jugadores.



Por ello, dentro de los estudios de análisis de datos existe una fase o etapa que es la segmentación de datos. En ella, se filtran cuáles son los datos, parámetros e indicadores que nos interesa valorar y, por consiguiente, nos permite eliminar todas aquellas muestras (en este caso jugadores) que no se adecuan a lo que busca el equipo.

Si un equipo ha encajado muchos goles durante una temporada, seguramente les interesará fichar a buenos defensas. Teniendo esta necesidad, ¿les interesará valorar a todos los jugadores de la liga o, incluso, a todos los defensas de la liga? La respuesta es evidente; no.

Utilizando las herramientas adecuadas, podrán filtrar la información de todos los jugadores y centrarse únicamente en aquellos futbolistas que tengan buenos datos de despejes de balones, robos de balón, pases acertados en su propio campo, etc. De este modo, la dirección técnica podrá centrarse en los jugadores que destaquen en estos aspectos y despreciar al resto. Después, habrá que valorar con criterios económicos y con otros criterios deportivos cual es el jugador más adecuado para fichar.

Si tuviéramos que resumir qué es la segmentación de datos, diríamos que es la fase o etapa en la que se establecen unos filtros para dividir el conjunto de muestras que tenemos en varios subgrupos o categorías para, de esta forma, trabajar sólo con las que nos interesan en un momento determinado.

Como finalización a este subapartado, se puede destacar que:

“La segmentación es una de las mejores maneras de innovar, porque en vez de encontrar un gran mercado permite identificar un segmento de mercado que nadie ha desarrollado.”

(Checa Cruz, 2019)

2.5.13. Raspado Web

Tal y como se comentará y analizará en apartados posteriores, en nuestro caso particular, no disponemos de la base de datos con la que se trabajará para desarrollar e implementar la herramienta de análisis en la gestión deportiva. Esto supone un problema en diversos aspectos, como la imposibilidad de realizar modificaciones en la estructura de los datos o la necesidad de dedicar un elevado número de horas a aprender y comprender el funcionamiento y la organización de la base. Así pues, se tendrá la obligación de obtener los datos de alguno de los portales web que actualmente ofrecen lo que se busca.

Con el fin de extraer los datos desde el portal web que se elija, se tendrá que emplear una técnica conocida como raspado web o, en inglés, *web scraping*.

Como definición, se podría decir que el raspado web es una técnica o procedimiento que se aplica para extraer u obtener datos e información de los sitios,



portales o páginas web. Esta técnica mantiene una fuerte relación con la indexación del sitio o página web.

Este concepto, también conocido como rastreo web, presenta una más que notable utilidad en multitud de situaciones, como sucederá en el caso práctico que se llevará a cabo más adelante. Sus beneficios y ventajas se pueden ver en multitud de sectores y son cada vez más las compañías que se están dando cuenta de esto e implementando estos procedimientos en sus tareas del día a día. Analizar cuáles son las tendencias en función de las búsquedas de los usuarios, estudiar la evolución de los portales de la competencia o tener acceso a determinada información son algunas de las muchas ventajas que ofrece el raspado web.

Sin embargo, y como conclusión a este subapartado, se debe mencionar que el raspado web no está exento de polémica y que son muchos los que ponen en tela de juicio la legalidad o no de esta técnica. En los casos más extremos, los usuarios que lleven a cabo un proceso de rastreo web se pueden arriesgar a incluso ser denunciados o demandados por parte de los propietarios del sitio web sobre el que se haya trabajado. Por ello, se recomienda intensamente conocer los términos legales y de uso de la página web que se quiera rastrear.

2.6. KPIs

Antes de dar una definición de lo que son los KPI, es imprescindible dejar claro que existen miles y que podemos “inventar” todos los que queramos. El caso del fútbol no es diferente, existen miles: Distancia recorrida, tiros a puerta, goles por partido, pases acertados, tarjetas amarillas, tarjetas rojas, pérdidas de balón, recuperaciones, faltas cometidas, minutos jugados, asistencias, despejes en el área y un larguísimo etcétera.

Y los anteriores eran sólo para jugadores, pero los hay también para los equipos (goles a favor, goles en contra, partidos ganados, partidos empatados, partidos perdidos, tarjetas amarillas, tarjetas rojas, posesión por partido, posesión media, goles en casa, goles como visitante, puntos sumados, puntos sumados en casa, puntos sumados como visitante, penales a favor, penales en contra, clasificación histórica, temporadas en primera división, temporadas en segunda división...), para los entrenadores (partidos ganados, empatados y perdidos, partidos como entrenador en primera división y en segunda división, número de veces amonestado, número de veces expulsado, porcentaje de victorias y derrotas, esquema táctico preferido, etc.) e incluso para las competiciones (número de ediciones disputadas, equipo más laureado, goles por temporada, máximo goleador, clasificación histórica, número de equipos participantes, jugadores con más partidos, países cuyos equipos pueden participar, federación de la que dependen, patrocinadores, canales por los que se retransmiten, número medio de espectadores...). Además, si prestamos atención, veremos cómo hay indicadores que son compartidos por varios grupos.



Seguramente, tras leer estos ejemplos ya podemos tener una idea más o menos clara de lo que son los KPIs. A continuación, se expondrán algunas definiciones y se estudiará su utilidad e importancia en el mundo del análisis de datos y el Big Data.

La primera definición que se muestra es la siguiente:

“KPI es un acrónimo formado por las iniciales de los términos: Key Performance Indicator. La traducción válida en castellano de este término es: indicador clave de desempeño o indicadores de gestión. Los KPIs son métricas que nos ayudan a identificar el rendimiento de una determinada acción o estrategia. Estas unidades de medida nos indican nuestro nivel de desempeño en base a los objetivos que hemos fijado con anterioridad.”

(Espinosa, 2016)

Una segunda definición es la siguiente:

El término KPI “hace referencia a una serie de métricas que se utilizan para sintetizar la información sobre la eficacia y productividad de las acciones que se lleven a cabo en un negocio con el fin de poder tomar decisiones y determinar aquellas que han sido más efectivas a la hora de cumplir con los objetivos marcados en un proceso o proyecto concreto. [...] El objetivo último de un KPI es ayudar a tomar mejores decisiones respecto al estado actual de un proceso, proyecto, estrategia o campaña y de esta forma, poder definir una línea de acción futura.”

(Porras Blanco, 2017)

Aunque estas definiciones van bastante enfocadas al ámbito empresarial y económico, son totalmente trasladables al ámbito de la gestión deportiva.

En cuanto a la utilidad que presentan, (Stoyanov Georgiev, 2019) destaca en su trabajo que las funciones y cometidos principales de un KPI son controlar uno o varios parámetros, informar de un modo entendible de los resultados, evaluar aquello que se está midiendo y, como consecuencia de lo anterior, ayudar a los responsables en la toma de decisiones. Y de esto que se acaba de exponer surge una de las características más definitorias de los indicadores clave de desempeño: la no existencia de una lista o repertorio prediseñado de KPIs, ya que éstos varían en cada organización, incluso entre aquellas empresas que forman un mismo sector.

“Los KPIs en sí mismo no son metas ni objetivos, sino que son una forma de medir el correcto funcionamiento de la organización para la consecución de sus objetivos. Proporcionan a la empresa información fundamental sobre el rendimiento del negocio, especialmente en periodos muy cortos. Sin el uso de estos indicadores tomaría mucho tiempo y esfuerzo realizar la recopilación y procesamiento de los datos para obtener información sobre el estado general



de la empresa; en esta medida se presentarían demoras en la identificación y tratamiento de los problemas que se puedan presentar.”

(Rivera Resina, 2018)

Uno de los puntos más importantes es el de las características que deben reunir todos los KPIs, sean del ámbito que sean. La mayoría de los autores están de acuerdo en que son las siguientes: Los KPIs deben ser:

- Medibles.
- Cuantificables.
- Específicos (o concretos): Deben centrarse en un único aspecto.
- Temporales: Deben poder medirse en el tiempo.
- Relevantes.
- Realistas.
- Objetivos.
- Coherentes.

Para terminar este apartado, se hablará de la importancia de saber qué indicadores debemos diseñar y emplear, pues de ello dependerá, en parte, el éxito de muchos proyectos. Como ya se ha comentado al principio de este apartado, también en el fútbol podemos diseñar e implementar cientos de indicadores de gestión que ayuden a los responsables técnicos de los equipos a tomar mejores decisiones a la hora de desarrollar la planificación deportiva de sus clubes. En el supuesto práctico se vivirá en primera persona la complejidad de este proceso y la importancia de emplear los criterios de evaluación adecuados si se quiere acertar con las decisiones tomadas.

Una reflexión importante en la que merece la pena detenerse es el hecho de que los KPIs pueden ser fundamentales para todo tipo de empresas, pero aún más si cabe en el caso de los equipos deportivos. Como por todos es sabido, la cantidad de dinero que mueve el deporte en general y el fútbol en particular es mareante para la mayoría de las personas; y es por ese motivo por lo que desembolsar una gran cantidad de dinero en el fichaje de un futbolista puede generar unos beneficios, tanto deportivos (victorias, títulos, trofeos individuales...) como económicos (ingresos por victorias, por una futura venta del jugador, por contratos de patrocinio, por ventas de camisetas y otros productos, etc.), inmensos, pero también puede ser un fiasco que hipoteque al club por un largo período de tiempo.

Para no caer en ese problema, se propone dar respuesta (o respuestas) a las siguientes cuestiones:

- *“¿Qué queremos medir?*
- *¿Por qué medimos este dato?*
- *¿Realiza el seguimiento de los resultados de uno de nuestros objetivos?*
- *¿Es un factor clave para la empresa?*



- ¿Quién es el responsable de supervisarlos?
- ¿Con qué periodicidad conviene supervisarlos?”

(Espinosa, 2016)

Con este punto se daría por completada la parte de Análisis de Datos y Big Data. A continuación, se desarrollará una parte de Inteligencia de Negocio o Inteligencia Empresarial (*Business Intelligence*), aunque se profundizará mucho menos que en ésta, ya que se verá de una forma más práctica en el supuesto que se planteará posteriormente.

Por último, antes de entrar en el caso práctico del trabajo, se hablará sobre el programa que se empleará, *Power BI*.



3. INTELIGENCIA DE NEGOCIO

3.1. ¿QUÉ ES LA INTELIGENCIA DE NEGOCIO?

Antes de embarcarnos en la búsqueda de una definición, lo más completa posible, de la inteligencia de negocio, conviene hablar de la relación entre ésta y el Big Data y el análisis de datos, que son los objetos de estudio de este trabajo, así como de su objetivo común.

“El Big Data nace a partir de la inteligencia de negocio (BI – Business Intelligence). El objetivo del BI es disponer de información con la que se pueda conocer el negocio, tomar decisiones y crear planes de acción con los que generar nuevos datos de interés para la organización. Para lo cual es necesario analizar todos los datos disponibles de la organización y transformar los mismos siguiendo un proceso básico que pasará por las [siguientes] fases:”
(Matesanz Niño, 2018)

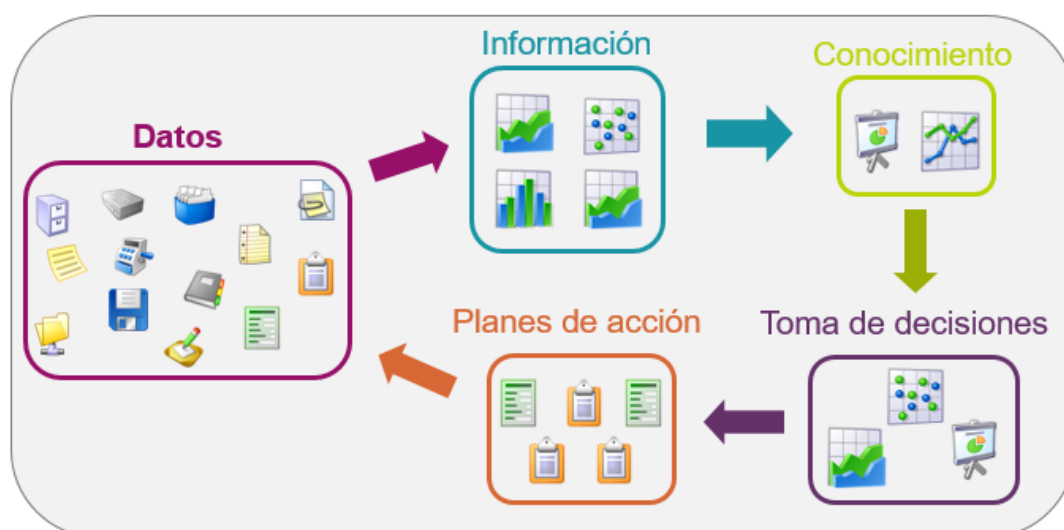


Ilustración 7. Fases del análisis de datos de la Inteligencia de Negocio. Fuente: (Matesanz Niño, 2018)

Como ya hemos comentado en puntos anteriores, la aparición de Internet (y más recientemente el Internet de las Cosas), las redes sociales y los dispositivos móviles han provocado un exponencial crecimiento de los datos generados y recopilados, por lo que ha sido necesario ir un paso más allá de la Inteligencia de Negocio. Ahí es donde tiene lugar la aparición del Big Data.

A continuación, trataremos de dar unas definiciones de lo que es la inteligencia de negocio o inteligencia empresarial, de modo que el concepto quede lo más claro posible.

“La consultora Gartner define business intelligence como: “BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un Data Warehouse), para descubrir



tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.”

(Checa Cruz, 2019)

“Podemos decir que la inteligencia empresarial o BI (business intelligence) se refiere a la ayuda que esta proporciona a una empresa para que dicha empresa sea capaz de “entenderse” mucho mejor y de manera más amplia sobre sí misma. A esto habría que sumarle todo lo positivo (aplicaciones, mejoras, habilidades) de las que saca fruto la empresa con todo esto. Esta datado en el año 1958 la primera vez que se utilizó el término, hoy por todos conocido, como inteligencia empresarial, “capacidad de aprehender las interrelaciones de los datos presentados de modo tal que se pueda orientar la acción hacia el objetivo deseado” de Hans Peter Luhn.”

(Martín Peña, 2016)

Atendiendo a la definición teórica propuesta por (Conesa Caralt & Curto Díaz, 2012) en su trabajo, se entiende, desde un punto de vista formal, por Inteligencia Empresarial o Inteligencia de Negocio al conjunto de metodologías, usos y utilidades, prácticas y competencias o capacidades enfocadas a la concepción y administración de los datos y la información que permiten a los directivos y a los empleados tomar mejores decisiones para un mayor beneficio de la organización y de las personas que la forman.

3.2. OBJETIVOS, CARACTERÍSTICAS Y VENTAJAS

En lo que respecta a la finalidad de la inteligencia de negocio, podemos decir lo siguiente:

El propósito general de la Inteligencia de Negocio y de sus herramientas asociadas es, según explica (Checa Cruz, 2019), amparar y avalar tanto a las grandes organizaciones como a las pequeñas y medianas empresas para mejorar su competitividad, allanando la transmisión de la información necesaria para una adecuada toma de decisiones.

Según comenta (Checa Cruz, 2019), las funciones que principalmente aporta la Inteligencia Empresarial en las compañías son: la confección y presentación de informes visuales fácilmente interpretables, el estudio detallado del rendimiento que presenta la empresa y los análisis de pronósticos futuros o predicción, todo ello extraído de los datos que se generan y recopilan permanentemente.

Si nos centramos en las características, debemos resaltar los siguientes aspectos:

“El Business Intelligence como concepto, reúne muchas características comunes a todas las herramientas de BI. Sin embargo, una es la que prevalece sobre las demás: la ayuda en la toma de las decisiones.”

(Stoyanov Georgiev, 2019)



Uno de esos aspectos es el que marca la diferencia con respecto al Big Data:

“El Business Intelligence ha conseguido actualizarse y analizar los datos recogidos, pero no ha logrado trabajar con la cantidad de datos que se generan actualmente. En la sociedad actual es necesario analizar los datos en tiempo real y esta herramienta no ha conseguido adaptarse a ello.”

(Checa Cruz, 2019)

Un sistema de inteligencia empresarial se compone de tres partes: la minería de datos, los mercados de datos y los almacenes de datos. La primera se estudiará en el siguiente apartado y las dos siguientes ya se vieron en la parte de análisis de datos y Big Data.

El principal inconveniente de la inteligencia de negocio es el que ya hemos resaltado: no es capaz de trabajar con los datos como lo hace un sistema actual de Big Data.

Si nos centramos en las ventajas, podemos mencionar los siguientes aspectos:

“Una de las grandes ventajas y beneficios del Business Intelligence es aglutinar todo ese montante de datos que recaen en las empresas (a mayor volumen de la empresa mayor será el de sus datos) y así poder correlacionarlos, manera en que podrán ser de utilidad. Por sí mismos los datos pueden no ser significativos, pero una vez que se ponen a disposición de la inteligencia empresarial adoptan formas que sí la tienen. Se suele decir, con toda la razón del mundo, que el BI no crea nuevos datos, sino que la parte en la que ayuda es en la relación y análisis de los mismos, ya que aun pareciendo disparatado o de difícil comprensión todo en las empresas está relacionado de alguna manera.”

(Martín Peña, 2016)

Otras ventajas que ofrece la Inteligencia de Negocio o Inteligencia Empresarial son las que expone (Martín Peña, 2016) en su trabajo, las cuales se enumeran y comentan a continuación:

- La posibilidad de visualizar los datos que contienen varias BBDD en una única consulta, de tal forma que exista la posibilidad de relacionar, por ejemplo, las ventas con algunos gastos determinados.
- Lograr adquirir una “foto” de los datos cosechados con respecto al tiempo pasado. Esto permite estudiar la evolución temporal del negocio, algo fundamental para detectar aspectos negativos.
- Otorgar a las empresas la capacidad para realizarse preguntas con vistas al futuro y obtener la respuesta (o respuestas) con datos de períodos anteriores.



Complementariamente a las ventajas expuestas hasta ahora, (Stoyanov Georgiev, 2019) analiza en su trabajo de fin de grado otros beneficios que la Inteligencia de Negocio puede reportar a las compañías que hagan uso de sus herramientas y procedimientos. Estos beneficios se comentan a continuación:

- El BI posibilita la integración de toda la información histórica recopilada y consolidada en términos de calidad.
- También ayuda a determinar, implantar y seguir la evolución de los indicadores clave de rendimiento o indicadores de gestión.
- Permite disponer de los datos y la información de forma permanentemente actualizada en la organización, tanto de un modo jerarquizado como detallado (por departamentos, proyectos, etc.).
- La Inteligencia Empresarial mejora notablemente la competitividad de la empresa con respecto al mercado en el que compete. Esto es debido a la adquisición de nuevas capacidades por parte de los sistemas y procedimientos de BI: mayor agilidad en la toma de decisiones, la ya comentada posibilidad de disponer de información actualizada...

3.3. MINERÍA DE DATOS

La minería de datos es uno de los conceptos más interesantes de los que se verán en este trabajo. Perfectamente se podría haber hablado de ello en la parte de Big Data y Análisis de Datos; sin embargo, he considerado más oportuno incluirlo en la sección de Inteligencia de Negocio (aunque como ya se ha comentado, ambas están íntimamente relacionadas) porque ésta es una parte muy relevante, posiblemente la que más, de cualquier herramienta de Inteligencia Empresarial.

En primer lugar, se van a estudiar algunas posibles definiciones de este concepto, de modo que se pueda comprender de una manera sencilla qué es y en qué consiste. Se podría empezar diciendo que:

“La minería de datos es un campo de la estadística y las ciencias de la computación dedicado a extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita que no es posible obtener mediante métodos estadísticos convencionales.”

(Moine et al., 2011)

Otra definición bastante completa es la que se muestra a continuación:

La “minería de datos es una tecnología que nos permite trabajar con grandes cantidades de bases de datos de manera automática o semiautomática, con el objetivo de encontrar patrones, tendencias o reglas que nos ayuden a comprender el comportamiento de los datos.”

(Rivera Resina, 2018)

Una tercera definición válida puede ser la siguiente:



“La minería de datos es una herramienta que nos ayuda a extraer conocimiento de los datos que tenemos almacenados para posteriormente a través de la utilización de determinadas técnicas y algoritmos puedan convertirse en información útil para la empresa.”

(Rivas González, 2019)

Por último, podríamos decir, de un modo resumido y sencillo, que la minería de datos es la aplicación al mundo empresarial o profesional de la inteligencia artificial y sus herramientas asociadas.

En cuanto al principal propósito que se persigue, se ha de destacar el siguiente aspecto:

“El objetivo principal de la minería de datos es filtrar de forma inteligente el conjunto de datos almacenados para, a partir de ahí, extraer la información relevante que estos datos almacenan.”

(Costalago Serrano, 2019)

Para concluir este apartado, es de obligada mención decir que, según relata (Rivas González, 2019) en su trabajo, en la actualidad existe lo que podríamos llamar un “estándar procedimental” denominado *CRISP-DM*. Dicho estándar permite fraccionar el proceso de la minería de datos en seis etapas o fases estructuradas. Dichas fases, según enumera (Rivas González, 2019) son las que citan a continuación:

- Interiorizar los requerimientos y los objetivos del proceso.
- Comprender los datos: Estructura, características, significados...
- Desarrollar el tratamiento de los datos para las próximas fases.
- Diseño en implementación del modelo.
- Evaluación de los resultados obtenidos.
- Despliegue: Presentación de informes, etc.

Sin embargo, como conclusión a este apartado, es necesario añadir que los avances y la rapidísima evolución que están experimentando el Big Data, el análisis de datos y la Inteligencia de Negocio en los últimos años, han provocado que algunos de estos estándares estén empezando a quedar obsoletos.

3.4. TABLEROS DE VISUALIZACIÓN

Si en el punto anterior se hablaba de la importancia de la minería de datos en la inteligencia de negocio, la parte de visualización no se queda atrás.

Normalmente, las conclusiones que se extraen tras trabajar con los datos y la información obtenida son complejas, ya que el volumen de información manejado puede haber sido muy grande. Si se tuvieran que explicar todas estas conclusiones con números o con un informe escrito, posiblemente, los responsables de tomar las decisiones no entenderían muchas de esas conclusiones y pasarían por alto otras cuantas, debido, entre otros aspectos, a la falta de conocimientos técnicos de los



responsables y a la extensa longitud que, con toda seguridad, tendría el informe en cuestión.

Por todo ello, dentro del mundo del Big Data y de la Inteligencia Empresarial, cada vez se le da mayor importancia a las herramientas y técnicas de visualización. Los tableros de visualización son los paneles en los que se exponen las conclusiones obtenidas de un modo gráfico e interactivo, mediante el uso de tablas, gráficas, diagramas (de barras, de área, de cajas y bigotes, de velas, de anillos, de sectores circulares...), histogramas, pirámides, mapas y un larguísimo etcétera.

Estos paneles o tableros de visualización permiten que, de un modo rápido y efectivo, el responsable de la toma de decisiones pueda ver las conclusiones más importantes, los puntos fuertes y las principales áreas de mejora de aquello que se está evaluando. Además, las nuevas aplicaciones y programas permiten que estos tableros sean interactivos, con lo que se pueden seleccionar muestras concretas para analizar su relación con el resto o para estudiar un caso determinado. Y, a mayores, se puede lograr que estos tableros trabajen en tiempo real o casi real, actualizándose permanentemente, con lo que las ventajas de estos se multiplican.

A continuación, se darán dos definiciones complementarias de lo que son los tableros de visualización, también conocidos por el término inglés, *dashboards*.

“Es una representación gráfica de los principales indicadores (KPI) que intervienen en la consecución de los objetivos de negocio, y que está orientada a la toma de decisiones para optimizar la estrategia de la empresa. Debe representar las KPIs necesarias, no por tener más indicadores será mejor la información, hay que saber hacer limpieza. Además, se deben presentar las KPIs de forma que sean relevantes, con un contexto. También, es importante una visualización que sea fácil de interpretar, que nos permita tomar decisiones de forma correcta y rápida.”

(Rivera Resina, 2018)

“Son la principal funcionalidad de las herramientas de BI. Los dashboard son elementos de las herramientas de BI que se utilizan a modo de panel de control en el que se expresan, generalmente de manera visual, todo tipo de información que se considera relevante. A la vez, se puede usar para representar los principales KPI definidos de la organización. La capacidad de personalización de estos dashboards es prácticamente ilimitada: desde colores, hasta disposición de las gráficas, generación de varias hojas (al estilo Excel) de dashboards, diseño, modificación de tamaños, etc. También suelen tener capacidad de convertirse en interactivos.”

(Stoyanov Georgiev, 2019)

En otros puntos del trabajo se han mostrado ilustraciones para comprender mejor de qué se estaba hablando, algo que en este caso tiene todo el sentido del mundo, al tratarse de una herramienta visual. Sin embargo, se omitirá esa



ilustración, ya que en la parte práctica se buscará llegar a un tablero de visualización del caso que se planteará. Por lo tanto, se mostrarán varias imágenes sobre ello, lo que permitirá aclarar aún mejor el concepto.

Con este apartado, se da por concluida la parte de Inteligencia de Negocio o Inteligencia Empresarial.

A continuación, se hablará sobre el programa *Power BI*.





4. POWER BI

Power BI es una herramienta (o, mejor dicho, un conjunto de herramientas) informática de inteligencia de negocio o inteligencia empresarial, desarrollada por *Microsoft* y cuya fecha de lanzamiento inicial fue el año 2014.

Está integrado dentro de *Office 365*, y sus principales componentes y características son los siguientes:

- La aplicación para escritorios de *Windows*, llamada *Power BI Desktop*.
- El servicio en línea, denominado *Power BI Service*, *PowerBI.com* o simplemente *Power BI*.
- Las aplicaciones para dispositivos móviles *Android* y *iOS* y para tabletas y teléfonos móviles de *Windows*.
- Puertas de enlace para conectarse a los orígenes de datos locales, para mantener los paneles e informes actualizados.
- Posibilidad de crear e insertar análisis e informes visuales, mediante *Power BI Embedded*.
- *Power BI Report Server*, que es la solución local del programa que permite crear informes en un momento determinado, con la posibilidad y flexibilidad de cambiar a la nube en otro momento.
- Posibilidad de crear objetos visuales personalizados (pudiendo emplear la biblioteca que ofrece el programa de visualizaciones de datos de código abierto) y compartirlos.

En resumen, se podría decir que *Power BI* ayuda a recopilar, administrar y analizar datos, los cuales pueden provenir de una gran variedad de orígenes distintos. Algunos de esos orígenes son los siguientes:

- *Excel*.
- Conjuntos de datos y flujos de datos de *Power BI*.
- Bases de datos de *Microsoft SQL Server*.
- Archivos de texto o *CSV*.
- Páginas web (que será la fuente que se empleará en el caso práctico de este trabajo).
- Fuentes *OData*.
- Y un larguísimo etcétera.

Una de las principales ventajas de *Power BI* es que permite utilizar gráficos y tablas fáciles de procesar y visualmente fáciles de interpretar. Gráficos de barras, gráficos de líneas, gráficos de áreas, gráficos de cascada, gráficos de dispersión, gráficos circulares y de anillos, mapas, embudos, medidores, tarjetas, tablas y matrices están entre su oferta de elementos de visualización.



Power BI permite manejar grandes cantidades de datos, presenta funciones integradas de aprendizaje automático y tiene una gran capacidad de personalización. Además, ofrece una interfaz intuitiva y fácil de utilizar.

Esta herramienta se integra con facilidad con otras plataformas de gestión empresarial y también ofrece a sus usuarios una elevada seguridad para sus datos e información.

Existe una versión gratuita y una versión de pago, la cual ofrece mayor capacidad de almacenamiento en la nube, la capacidad de compartir informes y algunas funciones y herramientas avanzadas.

Evidentemente, este apartado podría haber sido bastante más detallado, pero su brevedad se debe a que este Trabajo de Fin de Grado no es un manual de uso de *Power BI* ni un trabajo de investigación sobre este programa. Sin embargo, para todos aquellos que deseen profundizar en el entorno de *Power BI*, se muestran seguidamente los enlaces web a la página oficial del programa, al blog oficial y a la página de descargas de los distintos componentes de *Power BI*:

Página oficial: <https://powerbi.microsoft.com/es-es/>

Blog oficial: <https://powerbi.microsoft.com/es-es/blog/>

Página de descargas: <https://powerbi.microsoft.com/es-es/downloads/>

4.1. OTRAS HERRAMIENTAS EXISTENTES

En este trabajo se utiliza, por diversas razones, el programa *Power BI*. Sin embargo, existen otras muchas aplicaciones y herramientas similares en el ámbito de la inteligencia empresarial y el Big Data. A continuación, se mencionarán algunas de ellas y se compararán muy brevemente con *Power BI*.

4.1.1. Pentaho

Pentaho es una plataforma de inteligencia de negocio que, tal y como se autodefine, está orientada a la solución y centrada en los procesos. Cuenta con una infraestructura de herramientas para desarrollar análisis e informes. Ofrece diversos productos para mejorar la capacidad de análisis y de toma de decisiones de las empresas.

A diferencia de *Power BI*, *Pentaho* es un conjunto de herramientas y programas libres. Además, no presenta dependencia de ningún sistema operativo, el repositorio de datos está basado en XML y su plataforma es Java.

Sus principales ventajas son las siguientes: No se necesita adquirir licencias para su uso, funciona en distintas plataformas y sistemas operativos, usa tecnologías estándar (como Java o XML), es fácil de instalar y configurar, cuenta con un potente motor de minería de datos, está enfocado al Big Data y a las bases de datos no relacionales y su comunidad es muy amplia, con las ventajas que ello reporta.



Como desventajas, y contrariamente al caso de *Power BI*, se puede citar que la documentación fiable es escasa y que las herramientas fueron desarrolladas por separado y empleando diferentes metodologías.

Para profundizar más sobre este programa, se recomienda visitar la página web oficial: <https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform.html>

4.1.2. Tableau

Tableau es una potente herramienta de visualización de datos, que abarca el área de la inteligencia empresarial. Su principal función es simplificar datos en bruto y darles un formato fácil de entender.

Permite acceder a datos y analizarlos y crear informes y compartirlos mediante el uso de funciones simples e intuitivas. Su objetivo es que los usuarios puedan ver, de una forma rápida, patrones en sus datos.

Centrándonos en sus características, se pueden destacar las siguientes: puede tomar datos de muchas y muy diferentes BBDD (al igual que *Power BI*), se pueden mantener conexiones en vivo con actualizaciones automáticas, permite colaborar entre usuarios y compartir informes de una forma segura, tiene diseños optimizados para los dispositivos móviles y permite integrar paneles, entre otros elementos, en las aplicaciones ya existentes.

A diferencia de *Power BI*, *Tableau* sólo dispone de versión de pago, aunque ofrece una prueba gratuita durante 14 días. Sin embargo, otorga a sus clientes, sin coste adicional, algunos recursos adicionales y un programa de soporte técnico.

De una forma bastante similar a *Power BI*, *Tableau* cuenta con una aplicación para escritorio, un servidor (a través del cual se puede colaborar y compartir información de forma segura) y un servicio en línea mediante alojamiento en la nube. Además, dispone de aplicación móvil y de una versión completamente gratuita para compartir datos de forma pública.

Para profundizar más sobre este programa, se recomienda visitar la página web oficial: <https://www.tableau.com/es-es>

4.1.3. IBM Watson Analytics

IBM Watson Analytics es un servicio, basado en la nube, de inteligencia artificial. Su función es permitir a las empresas descubrir, analizar y visualizar, de una forma inteligente, datos e información relevante para descubrir patrones o tendencias y para ayudar a la toma de decisiones.

Dos de los principales puntos fuertes de *IBM Watson Analytics* son el análisis predictivo y la narración visual. Abarca muchas categorías, como ventas, operaciones, finanzas, atención al cliente, recursos humanos, etc. En todas ellas permite analizar y visualizar datos de forma interactiva, buscar de forma automática



patrones de comportamiento y analizar las redes sociales. A mayores, permite la creación de cuadros de mando.

Dentro del análisis deportivo, éste es menos empleado que los dos anteriores, que junto con *Power BI* son los más demandados en la actualidad.

Para profundizar más sobre este programa, se recomienda visitar la página web oficial: <https://www.ibm.com/es-es/watson>

4.1.4. Apache Spark

El último programa o herramienta que se verá es *Apache Spark*, que es un software de código abierto, desarrollado por la Universidad de California y donado posteriormente a la organización sin ánimo de lucro *Apache Software Foundation*, la cual lo gestiona y mantiene. Fue lanzado en mayo de 2014. Es la evolución natural de *Hadoop*.

Se puede definir como una solución Big Data orientada al procesamiento de datos. Permite el procesamiento de datos estructurados y de grafos. Se integra muy bien con otras herramientas de Big Data y análisis de datos.

Para profundizar más sobre este programa, se recomienda visitar la página web oficial: <https://spark.apache.org/>

Aunque no se analizan en este trabajo, resulta interesante mencionar algunas otras herramientas existentes, como *Qlik*, *SAS* o *SAP Predictive Analysis*.



5. CASO PRÁCTICO

5.1. PRESENTACIÓN DEL SUPUESTO

5.1.1. ¿Quién es el cliente?

Lo primero que se debe hacer a la hora de embarcarse en un proyecto o una investigación es documentarse acerca del cliente para el que se realiza. Éste puede ser una persona o un conjunto de personas, una empresa privada o una empresa pública, una organización sin ánimo de lucro o una ONG, etc. Además, podemos llevar a cabo un proyecto por cuenta propia.

En el supuesto que se planteará a continuación, el cliente será el Real Valladolid Club de Fútbol S.A.D., comúnmente conocido como Real Valladolid CF o, simplemente, Real Valladolid.

El Real Valladolid Club de Fútbol S.A.D. (en adelante, Real Valladolid), es un club de fútbol español situado en la ciudad de Valladolid, en la comunidad autónoma de Castilla y León. Actualmente juega en el Campeonato Nacional de Liga de Primera División, también denominado Primera División de España, torneo organizado por la Liga de Fútbol Profesional (LFP) y que también es conocido como La Liga o La Liga Santander, por motivos de patrocinio. Es la principal competición de fútbol del país. Además, ocupa el primer puesto según el coeficiente UEFA, por lo que es considerada por muchos como la mejor liga del mundo.

La Primera División de España cuenta con 20 equipos (38 jornadas por temporada), de los cuales los cuatro primeros clasificados obtienen plaza para jugar la Liga de Campeones de la UEFA (también conocida por su anterior nombre, Copa de Europa). El quinto y el sexto clasificado obtienen plaza para jugar la Liga Europa de la UEFA (y también el séptimo, en caso de que el ganador de la Copa del Rey esté entre los 6 primeros clasificados). Los 3 peores clasificados descienden al finalizar la competición a la Segunda División de España, de la cual también ascienden 3 equipos cada temporada a Primera División. La temporada 2020-2021 será la edición número 90 del torneo. El vigente campeón es el Real Madrid CF. A lo largo de la historia del torneo, únicamente nueve clubes han logrado ganarlo: Real Madrid CF, FC Barcelona, Atlético de Madrid, Athletic Club de Bilbao, Valencia CF, Real Sociedad de Fútbol, Real Betis Balompié, Sevilla FC y Real Club Deportivo de La Coruña. El club más laureado es el Real Madrid CF con 34 títulos.

Desde el 10 de abril del 2018, el entrenador del primer equipo del Real Valladolid es Sergio González Soriano. Sergio González tiene 43 años y es original de Hospitalet de Llobregat, en la provincia de Barcelona. Como jugador, desarrolló la mayor parte de su carrera entre el Real Club Deportivo Espanyol de Barcelona y el Real Club Deportivo de La Coruña. Además, ha sido internacional con la selección española de fútbol en 11 ocasiones y participó en la Copa Mundial de Fútbol de 2002. Ocupaba la posición de centrocampista.



Como entrenador, ha dirigido a las categorías inferiores del Real Club Deportivo Espanyol de Barcelona en los años 2013 y 2014 y al primer equipo entre 2014 y 2015. También es el seleccionador de la selección autonómica de fútbol de Cataluña, la cual no puede disputar torneos internacionales, cargo que ocupa desde el año 2016. En las dos temporadas y media que lleva como entrenador del Real Valladolid ha dirigido al equipo blanquvioleta en 95 encuentros, de los cuales ha ganado 31, ha empatado 29 y ha perdido 35.

Por su parte, el propietario del Real Valladolid es Ronaldo Luís Nazário de Lima, el cual también ejerce las funciones de presidente del club. Ostenta dichos cargos desde el 3 de septiembre del 2018. Ronaldo Luís Nazário de Lima es un exfutbolista brasileño, nacionalizado español, conocido común y deportivamente como Ronaldo. Tiene 43 años.

Como jugador de fútbol, militó en las filas del Cruzeiro Esporte Clube, PSV Eindhoven, FC Barcelona, FC Internazionale Milano, Real Madrid CF, A.C. Milán y Sport Club Corinthians. Además, ha disputado 105 partidos con la selección brasileña de fútbol, en los que ha anotado 67 goles. Es considerado por muchos como el mejor delantero centro de la historia del fútbol. Entre sus mayores logros están dos mundiales de fútbol ganados con Brasil (Estados Unidos 1994 y Corea del Sur-Japón 2002), una Copa Confederaciones con Brasil (Arabia Saudí 1997) y dos Copas América con Brasil (Bolivia 1997 y Paraguay 1999). A nivel individual, entre otros muchísimos reconocimientos, ha sido galardonado tres veces con el premio Jugador Mundial de la FIFA en los años 1996, 1997 y 2002. Ha ganado el Balón de Oro en dos ocasiones, 1997 y 2002, y la Bota de Oro en la temporada 1996/1997. Fue Balón de Oro de la Copa Mundial de Fútbol de 1998 y es el mejor jugador de la historia del fútbol para los votantes de *Goal.com*.

Desde el año 2003, Ronaldo también es embajador de las Naciones Unidas, del programa para el desarrollo en la lucha contra la pobreza. Y desde el año 2016, es embajador mundial del Real Madrid CF.

En la directiva del Real Valladolid también están Matthieu Fenaert, como Consejero Delegado, Carlos Suárez, como Presidente de Honor y consejero, y Javier Ferrero Muñoz, como secretario no consejero del Consejo de Administración.

El Real Valladolid se fundó el 20 de junio de 1928, fruto de la unión de dos equipos de la ciudad, el Club Deportivo Español y la Real Unión Deportiva. El blanco y el violeta son los colores que luce el equipo en su equipación titular desde su fundación. En el año 1984, el Real Valladolid ganó la Copa de la Liga (torneo ya desaparecido), que es el único trofeo oficial que ha logrado hasta la fecha. En dos ocasiones ha sido subcampeón de la Copa del Rey (temporadas 1949/1950 y 1988/1989), ha participado en la Copa de la UEFA (predecesora de la actual Liga Europa) en las temporadas 1984/1985 y 1997/1998 y en la también extinta Recopa de la UEFA durante la temporada 1989/1990.



Habiendo disputado 45 temporadas en la Primera División, es el club más importante, por historia y palmarés, de la comunidad autónoma de Castilla y León. En la clasificación histórica de La Liga, ocupa la decimotercera posición, justo detrás del Real Club Deportivo de La Coruña y por delante del Real Racing Club de Santander.

Actualmente, el Real Valladolid disputa sus partidos como local en el Nuevo Estadio José Zorrilla. El estadio fue inaugurado el 20 de febrero del año 1982 y es de titularidad municipal. Tiene un aforo de 27.618 localidades. En la temporada 2019/2020 el Real Valladolid batió su récord de abonados con 22.212. La ciudad deportiva del Real Valladolid son los Campos Anexos al Estadio José Zorrilla, instalación que fue inaugurada en la temporada 1988/1989.

Con la información presentada, ya nos podemos hacer una idea más o menos clara de la empresa sobre la que versará el supuesto práctico. Sabemos en qué sector está y quien es su propietario y presidente. Dentro del sector, el deporte y, más concretamente, el fútbol, sabemos qué competiciones juega y quien es su entrenador. A mayores, disponemos de unas nociones básicas de su historia y de su situación actual.

El siguiente paso, y posiblemente el más importante de todo el proyecto, es conocer cuáles son los requisitos que pide el cliente. En otras palabras, debemos saber qué quiere.

5.1.2. ¿Qué quiere el cliente?

Como se acaba de mencionar, conocer con precisión cuales son las exigencias o requisitos del cliente es fundamental para llevar a buen puerto el proyecto. Por todos es sabido que, a la hora de entregar los resultados de un proyecto, muchas veces el cliente se muestra disconforme porque lo que se ha hecho no era lo que él pedía. Puede parecer algo extraño, pero es mucho más frecuente de lo que nos podemos llegar a imaginar. Una duda que se quedó sin preguntar, una falta de entendimiento al comentar algún aspecto, un punto del pliego de condiciones que se pasó por alto, algo que al cliente se le olvidó decir y que dio por hecho que sabíamos... son algunas de las causas que pueden facilitar que no se haga exactamente lo que el cliente solicitó.

En el supuesto que se plantea en este trabajo, la situación es la siguiente:

El Real Valladolid quedó en la temporada 2019/2020 en la decimotercera posición del Campeonato Nacional de Liga de Primera División, con 42 puntos. Ganó 9 partidos, empató 15 y perdió en 14 ocasiones. Justo por delante quedó el Levante UD, con 49 puntos. Por su parte, precedió a la SD Éibar, que también finalizó el campeonato con 42 puntos. La línea de salvación la marcó el Real Club Celta de Vigo, con 37 puntos. El CD Leganés quedó en decimotercera posición con 36 puntos. Si miramos hacia arriba en la clasificación, el último billete para disputar competición europea lo consiguió el Granada CF, séptimo clasificado con 56 puntos.



Centrándonos de nuevo en el Real Valladolid, consiguió anotar 32 goles y, por el contrario, encajó 43, lo que hace una diferencia de -11 goles. El jugador que más partidos jugó fue Óscar Plano, con 36 encuentros disputados, y el que más veces fue titular Jordi Masip, con 35. Éste último, además, fue el que más minutos disputó a lo largo de toda la liga, 3150 minutos.

El jugador que más tarjetas amarillas recibió fue Rubén Alcaraz, con 10 amonestaciones. Ningún jugador recibió una tarjeta roja en toda la temporada. Por otro lado, el máximo goleador del equipo fue Sergi Guardiola, con 8 goles. Por último, los jugadores que más partidos ganaron fueron Óscar Plano, Sergi Guardiola, Javi Moyano y Enes Ünal, que participaron en las nueve victorias del conjunto blanquvioleta.

Conociendo todos estos datos, el Real Valladolid desea llevar a cabo la planificación deportiva para la próxima temporada, en la que aspira a mejorar puestos en la clasificación, alejándose aún más de las posiciones de descenso y acercándose a las plazas que dan acceso a jugar la Liga Europa de la UEFA. Plazas que espera lograr en un plazo menor de 5 años.

Para desarrollar dicha planificación, en primer lugar, desea conocer el rendimiento de sus jugadores. El club quiere que, viendo el rendimiento global o general de la plantilla, definamos unos indicadores con los que evaluar el rendimiento de cada futbolista del equipo comparado con el resto (lógicamente, habrá aspectos que serán medidos por posiciones). Hecho ese análisis, debemos presentar unos resultados que permitan a la dirección deportiva del equipo decidir qué jugadores deben continuar la próxima temporada (decidiendo también a quién se debe renovar si acaba contrato este año), qué jugadores deben ser vendidos o cedidos y qué posiciones se deben reforzar más urgentemente.

Haciendo hincapié en esto último, el segundo paso consiste en, una vez se ha decidido qué posiciones se deben reforzar, buscar entre el resto de los equipos españoles de primera, segunda y segunda división B a jugadores que pudieran, por sus características y estadísticas, ser fichados. Puesto que la importancia de este trabajo radica en operar con un volumen de datos más o menos grande, dentro de las limitaciones que se irán comentando según se desarrolla la herramienta, también se trabajará con las estadísticas de los jugadores de los equipos de las primeras divisiones de Inglaterra, Francia, Alemania, Italia, Portugal y Holanda.

Se deberán tener en cuenta aspectos como la división en la que juega, puesto que no tiene el mismo mérito anotar un gol en primera división que hacerlo en segunda B, por ejemplo. También cabe mencionar que, aunque el criterio económico no formará parte de la herramienta de análisis en la gestión deportiva, a la hora de presentar los resultados al Real Valladolid, no se recomendará el fichaje de ningún jugador cuyo valor de mercado sea superior a los 10 millones de euros. Esta decisión se ha tomado para evitar que, tras obtener los resultados de la herramienta, ésta “diga” que conviene fichar a Leo Messi, Sergio Ramos o Jan Oblak; jugadores que, a



todas luces, están fuera del alcance de un equipo cuyo objetivo es quedar en torno a la posición 10 u 11 para la próxima temporada.

A lo largo del proceso surgirán otros problemas u otras cuestiones que se comentarán y resolverán según vayan surgiendo a lo largo del supuesto práctico.

Una vez conocemos qué es lo que quiere el cliente, es decir, el Real Valladolid, llega el momento de decidir cómo vamos a llevar a cabo la herramienta de análisis en la gestión deportiva.

5.1.3. ¿Cómo se va a desarrollar la herramienta de análisis en la gestión deportiva?

Lo primero que se debe decidir es cuál va a ser la fuente de la que se van a tomar los datos de los futbolistas.

Por su organización, facilidad para “moverse” por la página web, claridad de la información, encontrarse en idioma español, cantidad de información proporcionada, fiabilidad y otra serie de aspectos, se ha decidido que la base de datos que se utilizará en este Trabajo de Fin de Grado sea *BD Fútbol*, a la cual se puede acceder a través del siguiente enlace:

<https://www.bdfutbol.com/es/index.html>

Con relación a las estadísticas que *BD Fútbol* recopila de cada jugador, a continuación, se enumerarán cuáles son los datos que se miden para cada jugador de cada equipo (por temporada):

- Dorsal.
- Nacionalidad (no se tendrá en cuenta en nuestro caso práctico).
- Nombre o apodo.
- Posición.
- Partidos que ha jugado.
- Partidos en los que ha sido titular.
- Partidos que ha disputado completos.
- Partidos en los que ha sido suplente.
- Partidos en los que ha sido sustituido.
- Partidos en los que, yendo convocado, no ha participado.
- Minutos jugados a lo largo de toda la liga.
- Tarjetas amarillas.
- Tarjetas rojas.
- Goles.
- Minutos necesarios para anotar un gol (en promedio).
- Goles que ha anotado en tiros penales.
- Goles en propia puerta.
- Goles encajados (en el caso de los porteros o guardametas).
- Partidos ganados.
- Partidos empatados.



- Partidos perdidos.

Estos son los datos que se utilizarán para implementar la herramienta de análisis en la gestión deportiva. Sin embargo, *BD Fútbol* ofrece aún más información de los jugadores, la cual en principio no será relevante para obtener los resultados que desea el Real Valladolid pero que, de manera puntual, podría ser consultada y tenida en cuenta a la hora de tomar las decisiones finales. Esa información es la que se muestra a continuación:

- Edad.
- Fecha de nacimiento.
- Lugar de nacimiento y provincia.
- Jugadores que han sido incorporados en la última temporada y su equipo de procedencia.
- Jugadores que salieron en la última temporada y su equipo de destino.

Además, si se accede a la ficha de cada jugador se puede encontrar la siguiente información:

- Altura.
- Peso.
- Trayectoria: Equipos en los que ha jugado desde su debut.
- Palmarés.
- Trayectoria a nivel de selección.
- Estadísticas a lo largo de su carrera profesional. Por ejemplo:
 - Temporadas en cada categoría.
 - Número de partidos en los que ha anotado al menos un gol.
 - Dobletes conseguidos.
 - Máximo número de jornadas consecutivas jugando.
 - Máximo número de jornadas consecutivas ganando.
 - Máximo número de jornadas consecutivas perdiendo.
 - Etcétera.

Con lo expuesto hasta ahora, ya podemos hacernos una idea de cuál es la base de datos que servirá como fuente para obtener las estadísticas de los jugadores. El siguiente paso consiste en definir cuáles son los indicadores y criterios que se van a emplear para evaluar el rendimiento de los futbolistas.

Como ya se ha comentado anteriormente, habrá algunos que serán válidos para todos los jugadores de la plantilla. Sin embargo, habrá otros que serán definidos para cada posición. Lógicamente, no tiene sentido evaluar a un portero por los goles de penal que anota o por los partidos en los que ha sido sustituido (ya que, por norma general, el portero que comienza un partido siempre lo termina, salvo que sea sustituido por una lesión). Tampoco tendría sentido medir el rendimiento de un defensa por los goles que anota pues, aunque pueda anotar muchos, su trabajo



consiste en garantizar la solidez defensiva del equipo. Sin embargo, eso no quita para que, posteriormente, se tuviera en cuenta de forma positiva si también aporta buenos datos en otras facetas del juego, como por ejemplo el ataque.

Los ejemplos aportados deberían servir para certificar la importancia que tiene la etapa de definición de los indicadores que se deberán emplear en la herramienta. A continuación, se enumerarán cuáles son los que se pueden utilizar en nuestro supuesto práctico (ahora sólo se nombrarán y, en los próximos apartados, se profundizará más en cuáles y cómo se implementará cada uno):

- Indicadores generales para todos los futbolistas:
 - Partidos jugados.
 - Partidos como titular.
 - Partidos disputados de forma completa.
 - Partidos en los que es sustituido y en los que es suplente, con respecto al número de encuentros en los que participa.
 - Minutos jugados.
 - Partidos ganados con respecto al número de encuentros en los que juega.
 - Partidos empatados con respecto al número de encuentros en los que juega.
 - Partidos perdidos con respecto al número de encuentros en los que juega.
- Indicadores específicos para los defensas (centrales y laterales):
 - Tarjetas amarillas recibidas, con respecto a los minutos jugados.
 - Tarjetas amarillas recibidas, con respecto al total de las tarjetas recibidas por el equipo.
 - Tarjetas rojas recibidas, con respecto a los minutos jugados.
 - Tarjetas rojas recibidas, con respecto al total de las tarjetas recibidas por el equipo.
 - Goles en propia puerta, con respecto a los minutos jugados.
- Indicadores específicos para los centrocampistas:
 - Partidos en los que parte como suplente, con respecto al número de partidos en los que participa.
 - Partidos en los que no participa, pese a ir convocado.
 - Tarjetas amarillas recibidas, con respecto a los minutos jugados.
 - Tarjetas rojas recibidas, con respecto a los minutos jugados.
 - Minutos necesarios para anotar un gol.
- Indicadores específicos para los delanteros:
 - Goles anotados, con respecto al total de goles anotados por el equipo.



- Minutos necesarios para anotar un gol.
- Goles anotados de tiro penal, con respecto al número total de penales anotados por el equipo.
- Partidos en los que parte como suplente, con respecto al número de partidos en los que participa.
- Partidos en los que no participa, pese a ir convocado.
- Indicadores específicos para los porteros o guardametas:
 - Goles encajados, con respecto a los minutos jugados.
 - Goles encajados, con respecto al total de los goles encajados por el equipo.

No se evaluarán las expulsiones debido a que el Real Valladolid no recibió ninguna tarjeta roja a lo largo de los 38 partidos de la liga en la temporada 2019/2020. Se podrán analizar en el caso de jugadores de otros equipos, llegado el momento de decidir qué fichajes se deben abordar. Ídem con los goles en propia puerta.

Una vez se ha explicado cómo se va a desarrollar la herramienta de análisis en la gestión deportiva, analizando los datos que se van a emplear y definiendo los indicadores con los que se evaluará el rendimiento de los jugadores; ya podemos embarcarnos en la implementación de la herramienta en *Power BI*.

5.2. IMPLEMENTACIÓN DE LA HERRAMIENTA EN POWER BI

5.2.1. Carga de las tablas en Power BI

En primer lugar, es necesario indicar que, posiblemente, éste haya sido el paso más complicado de todo el supuesto práctico. Cargar una tabla en *Power BI*, desde una página web, es muy simple y sencillo, como se verá a continuación. Sin embargo, lograr cargar muchas tablas con el mismo formato, pero que se encuentran en enlaces diferentes, de forma conjunta, ha requerido de un complejo proceso de investigación y pruebas.

En el supuesto planteado, nos interesa obtener las tablas con las estadísticas de todos los jugadores de cada uno de los equipos de primera, segunda y segunda división B de España y de las primeras divisiones de Inglaterra, Francia, Alemania, Italia, Portugal y Holanda. Como cabría esperar, cada tabla se encuentra en un enlace diferente, el correspondiente a cada uno de los equipos. Por lo tanto, si se decidieran cargar una a una las tablas, introduciendo la dirección URL correspondiente en cada caso, tendríamos que copiar el enlace de cada uno de los veinte equipos de primera división, los veintidós equipos de segunda división y los veinte equipos de cada uno de los cuatro grupos de segunda división B, lo que hace un total de 122 equipos, es decir, 122 enlaces web. Y eso sólo sería para una temporada; si quisiéramos cargar la temporada que acaba de terminar y la anterior serían 244 enlaces, si quisiéramos cargar las tres últimas temporadas serían 366 enlaces, y así sucesivamente. Y esto sólo para los equipos españoles, si incluimos los clubes de los países que se han



indicado, la cantidad de direcciones URL que habría que añadir sería aún muchísimo mayor.

Queda claro que ese método sería inasumible, por su lentitud y su ineficiencia entre otras razones. Por lo tanto, se debe plantear una solución mejor. Obtener un método más óptimo para cargar las tablas en *Power BI* pasa por responder a la siguiente pregunta: ¿Existe en las direcciones URL algún elemento que se pueda parametrizar o que permita, de alguna forma, iterar a *Power BI*?

La respuesta es sí. Estudiaremos el caso con un ejemplo, para entender mejor qué es lo que se quiere hacer. A continuación, se muestra el enlace en el que está registrada la tabla de estadísticas de los jugadores del Real Valladolid:

<https://www.bdfutbol.com/es/t/t2019-2037.html?t=lista>

Para poder apreciar cuáles son las partes comunes de las direcciones URL, cuáles son las partes que cambian y cuáles de estas últimas se podrían parametrizar, se mostrará el enlace homólogo de otro equipo. Por ejemplo, el Real Madrid CF:

<https://www.bdfutbol.com/es/t/t2019-202.html?t=lista>

Tal y como se puede ver, el enlace es exactamente igual excepto que, después de la temporada (expresada como “2019-20”), en el caso del enlace del Real Valladolid aparece “37” y en el caso del Real Madrid CF aparece “2”. Estos números, por lo tanto, son los que hacen referencia al código del equipo. Y, evidentemente, éste debe ser el elemento por parametrizar.

Sin embargo, esto no es todo. Tal y como se ha mencionado anteriormente, puede interesarnos obtener los datos de dos o más temporadas. Para ver qué cambia en ese caso, se volverá a tomar el enlace del Real Valladolid y se comparará con el de la temporada anterior, que es el siguiente:

<https://www.bdfutbol.com/es/t/t2018-1937.html?t=lista>

En este caso, se mantiene el código del equipo y lo único que cambia es el identificador de la temporada, “2018-19” en lugar de “2019-20”.

A través de diferentes fuentes, se ha llegado a la conclusión de que la mejor forma, o una de las mejores, de parametrizar algo en *Power BI* es creando una función. Lo que debe hacer esa función es pasar variables a *Power BI*, que vayan modificando la dirección URL de la que se desea extraer información. De este modo, el programa irá leyendo todas las direcciones web e irá importando todas las tablas. Dicho de otra manera, se debe provocar una iteración que permita leer todos los enlaces web donde vienen las tablas de todos los equipos. Además, se debe conseguir que todas esas tablas se muestren como una sola. Para ello, haremos lo siguiente:



En primer lugar, se carga una tabla (la de cualquier equipo valdría) como se haría de forma normal. Ésta servirá a *Power BI* como referencia de cómo deben ser todas las demás tablas que queremos obtener. Cargaremos, por ejemplo, la tabla de estadísticas de los jugadores del Real Valladolid; la que, como ya se ha comentado anteriormente, se encuentra en el enlace:

<https://www.bdfutbol.com/es/t/t2019-2037.html?t=lista>

Para cargarla, seguimos los siguientes pasos:

Inicio → *Obtener Datos* → *Web* → *Dirección URL* → *Aceptar*

Una vez hecho esto, el programa encuentra en el enlace web proporcionado todas las tablas que haya, permitiendo una visualización de las éstas. En ese momento se selecciona la tabla, o tablas, que deseemos y se pulsa “*Cargar*”.

En segundo lugar, se debe abrir el *Editor de Power Query*, donde se pueden hacer multitud de funciones, como especificar datos, administrar parámetros, quitar columnas y/o filas, dividir columnas, agrupar filas, combinar consultas y otras muchas posibilidades que el programa nos ofrece. En nuestro caso práctico, lo adecuado es abrir el “*Editor Avanzado*”. En él, veremos el código en lenguaje M que emplea el programa. Por suerte, no es necesario saber programar en profundidad ni ser experto en el lenguaje para poder modificar algunas cosas. Lo que vemos al abrir el *Editor Avanzado* es lo siguiente:



Ilustración 8. Pantalla del Editor Avanzado de Power Query. Fuente: Elaboración Propia

Lo tercero que se debe hacer es modificar el código que aparece. Concretamente, es necesario cambiar la dirección URL que está escrita por una función que itere cada código de cada equipo, para que el programa pueda leer todos los enlaces, y no sólo uno.



Para ello, se debe indicar que se va a pasar un código, el cual será un valor numérico, y que éste permitirá leer un enlace web que va a devolver una tabla como conjunto de resultados. Esto se especificará en la primera línea, antes de la función. Esa función será “*Number.ToText*”, la cual permite convertir en una cadena legible la variable que se le pase, en este caso el código de cada equipo del que queramos obtener las estadísticas de sus jugadores; código al que se denominará con el nombre “*codigo_equipo*”. Tras modificar todo esto, el editor avanzado muestra lo siguiente:

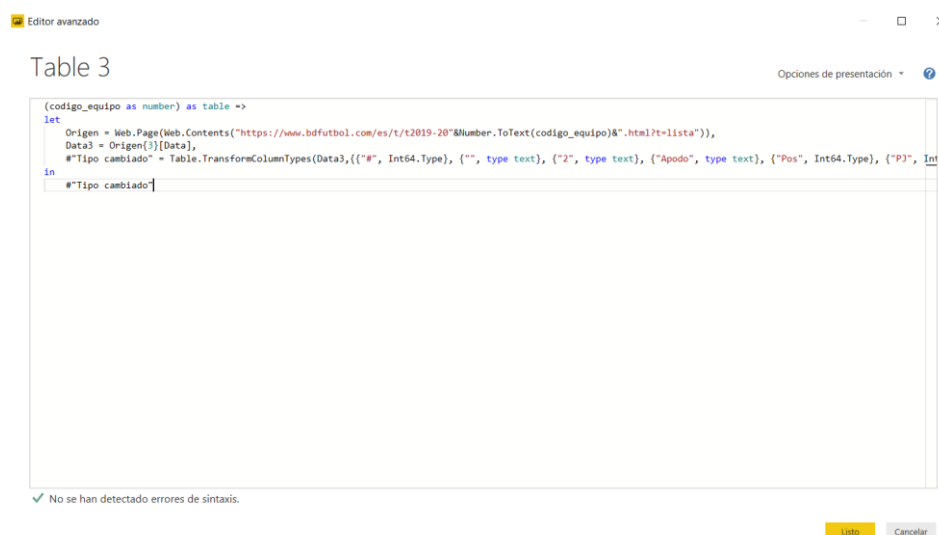


Ilustración 9. Pantalla del Editor Avanzado de Power Query una vez se han llevado a cabo todas las modificaciones indicadas. Fuente: Elaboración Propia

Es imprescindible destacar la importancia de comprobar que aparezca el mensaje “*No se han detectado errores de sintaxis*”, que se encuentra en la esquina inferior izquierda del editor, antes de pulsar “*Listo*”. Este mensaje nos indica que el código no presenta errores y que, por lo tanto, está listo para ser utilizado. En caso contrario, en lugar de ese mensaje aparecerá otro indicando el error concreto que hay y que debe ser subsanado.

En este punto, el programa nos permite escribir el parámetro (el código del equipo) que queramos y nos carga la tabla correspondiente al número indicado, sin necesidad de introducir la dirección URL. Es un gran avance, pero aún no es lo que buscamos exactamente. Para solucionar esto, crearemos una tabla en la que escribiremos los códigos de todos los equipos (se pondrán los números del 1 al 6999, aunque se sabe que hay muchos menos equipos). Esto lo haremos de la siguiente forma:

Inicio → *Obtener Datos* → *Consulta en blanco*

Una vez se haga esto, aparecerá una barra al lado del símbolo de función, donde deberemos solicitar que genere una columna con los números del 1 al 6999; lo que, tras haber investigado cómo hacerlo, se escribe así: = {1..6999}



Seguidamente, se debe transformar dicha columna en una tabla, para lo cual se debe pulsar en “*A la tabla*”. Se podrán, además, hacer otras cosas, como cambiar el nombre de la columna (en el supuesto planteado se pondrá “*codigo_equipo*”) o indicar el tipo de valores que contiene, que en este caso son números enteros.

Recordemos que lo que se pretende conseguir es una función que itere el valor de un parámetro, en este caso el código del equipo, con cargo a los valores de la tabla que se ha creado. Para conseguir este objetivo, se debe agregar una columna personalizada. La pantalla que aparece al principio es la siguiente:

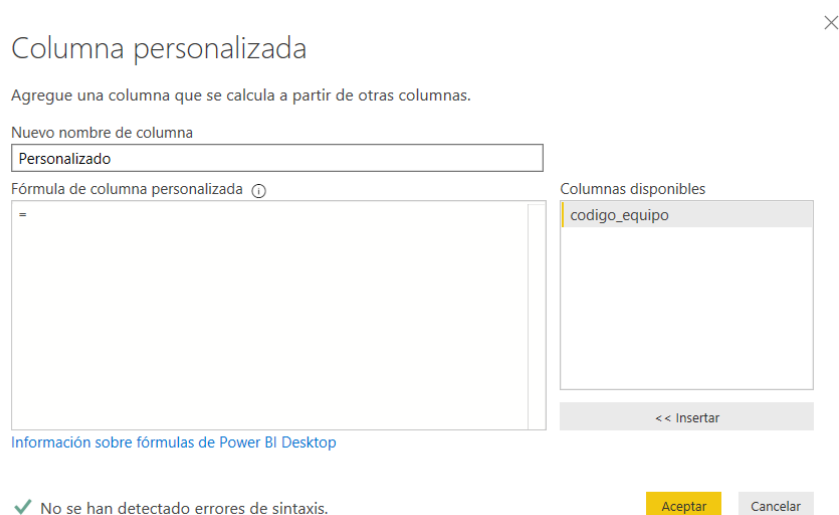


Ilustración 10. Pantalla donde definir la columna personalizada a agregar. Fuente: Elaboración Propia

En la pantalla de “*Columna Personalizada*” se debe indicar, como podemos ver, el nombre de la columna y la fórmula de ésta. En el supuesto que se plantea, la fórmula debe pasar la columna donde aparecen registrados los códigos de los equipos y la función creada previamente. Por supuesto, debe aparecer el mensaje de ausencia de errores antes de pulsar sobre “*Aceptar*”. Sabiendo que en *BD Fútbol* la tabla donde vienen las estadísticas de los jugadores de cada equipo se llama “*Table 3*”, el resultado debe ser el que se muestra en la siguiente ilustración:



Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna
datos_jugadores

Fórmula de columna personalizada ⓘ
= #"Table 3"[[[codigo_equipo]]]

Columnas disponibles
codigo_equipo

<< Insertar

✓ No se han detectado errores de sintaxis. Aceptar Cancelar

[Información sobre fórmulas de Power BI Desktop](#)

Ilustración 11. Pantalla de la columna personalizada, con ésta ya definida. Fuente: Elaboración Propia

Power BI nos requerirá información sobre la privacidad de los datos, pero da la opción de ignorar las comprobaciones de los niveles de privacidad. En este caso, se ignorará.

Si todo se ha hecho de forma adecuada, al lado de la columna “*codigo_equipo*” aparecerá otra columna llamada “*datos_jugadores*”. Esta columna podrá contener dos palabras: “*Table*” o “*Error*”.

La primera palabra quiere decir que *Power BI* ha encontrado una tabla en la dirección URL correspondiente al código de la columna “*codigo_equipo*”. Es decir, ese código corresponde a un equipo que en la temporada 2019/2020 militaba en la primera, la segunda o la segunda división B de España o en alguna de las primeras divisiones de Inglaterra, Francia, Alemania, Italia, Portugal u Holanda. Posteriormente se explicará qué conjunto de códigos corresponden a cada país, en función de lo establecido por los programadores de *BD Fútbol*.

El segundo término que nos podemos encontrar es “*Error*”. El significado, en este caso, es que no existe una dirección URL llamada así, bien porque el código corresponda a un equipo que ha desaparecido o a un equipo que no milita en ninguna de las categorías que se han indicado.

En la siguiente ilustración se puede ver el resultado obtenido:

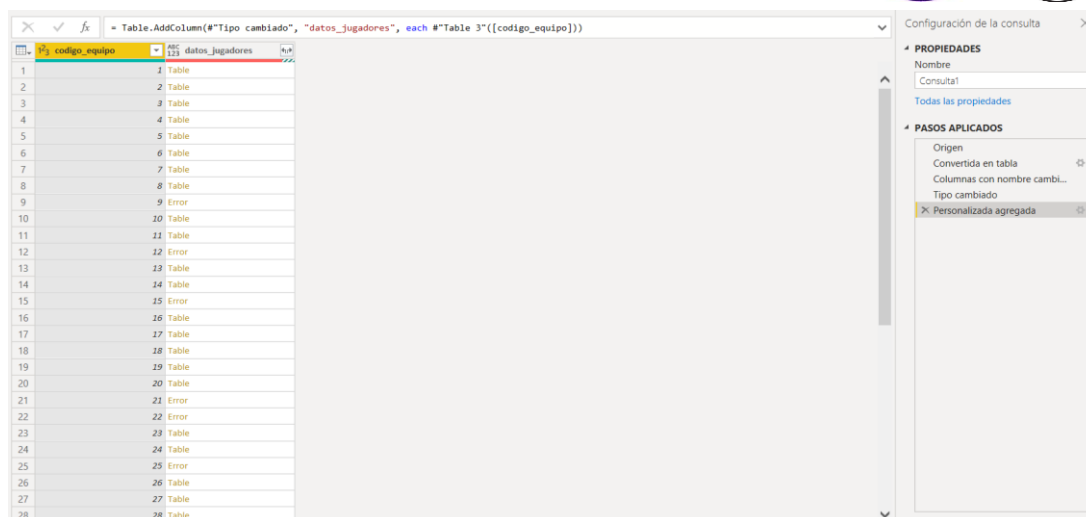


Ilustración 12. Pantalla del Editor de Power Query, una vez creada la columna donde aparece la tabla de cada equipo. Fuente: Elaboración Propia

Como se puede apreciar en la ilustración, las palabras “Table” y “Error” aparecen como enlaces. Esto quiere decir que, si pinchamos en una de ellas, se mostrará por pantalla la tabla que se haya seleccionado (o, en su defecto, un mensaje de error con el motivo por el cual no se ha encontrado ninguna tabla). Pero se ha de recordar que lo que se pretende es generar una “macro-tabla” en la que aparezcan las estadísticas de todos los jugadores de todos los equipos. Para lograr eso se realizarán dos pasos.

En primer lugar, se eliminarán todas las filas que no tengan una tabla asociada; es decir, se eliminarán todas las filas donde aparezca “Error”. *Power BI* permite hacer esto de una forma muy sencilla. Al apoyarnos en la barra roja que está debajo del nombre de la columna, donde muestra el porcentaje de filas con errores, aparece el mensaje “Quitar Errores”. Pulsamos sobre él y el programa elimina todas esas filas automáticamente. El proceso durará varios minutos. Una vez se ha hecho esto, el segundo paso consiste en pulsar sobre el icono con dos flechas que aparece al lado del nombre de la columna “datos_jugadores”. Ese icono permite expandir las columnas que queramos de las tablas, permitiendo elegir cuales son las que queremos mostrar. En nuestro caso, expandiremos todas las columnas excepto las dos que aparecen en blanco, que son las que corresponden, en la web de *BD Fútbol*, con la imagen del rostro del jugador y con la imagen de la bandera del país del jugador.

Más adelante trataremos de obtener la imagen de cada jugador, para dar un aspecto visualmente más agradable al comparador de jugadores. En el subapartado correspondiente se explicará cómo.

Si todo se ha hecho de forma correcta, la “macro-tabla” donde aparecen las estadísticas de los jugadores de todos los equipos ya estaría creada. Su aspecto es el que se muestra a continuación:



P3	codigo_equipo	ABC 123 #	ABC 123 Apodo	ABC 123 Pos	ABC 123 PJ	ABC 123 PT	ABC 123 PC
24	1	46	Reis	6	0	0	0
25	1	8	Arthur	6	21	14	14
26	1	19	Aleñá	6	4	2	2
27	1	7	Coutinho	6	0	0	0
28	1	22	Vidal	6	33	16	16
29	1	28	Riqui Puig	6	11	5	5
30	1	5	Busquets	6	33	29	29
31	1	4	Rakitić	6	31	16	16
32	1	29	Abel Ruiz	9	0	0	0
33	1	19	Braithwaite	9	11	4	4
34	1	9	Luis Suárez	9	28	22	22
35	1	11	Dembélé	10	5	3	3
36	1	41	Manaj	10	0	0	0
37	1	27	Carles Pérez	10	10	5	5
38	1	38	Kike Saverio	10	0	0	0
39	1	31	Ansu Fati	10	24	11	11
40	1	17	Griezmann	10	35	31	31
41	1	10	Messi	10	33	32	32
42	1	45	Konrad	10	0	0	0
43	2	13	Courtois	1	34	34	34
44	2	1	Areola	1	4	4	4
45	2	46	Luis	1	0	0	0
46	2	1	Keylor Navas	1	0	0	0
47	2	26	Altube	1	0	0	0
48	2	19	Odrizola	2	4	4	4
49	2	2	Carvajal	2	31	31	31
50	2	23	Mendy	3	25	19	19
51							

Ilustración 13. Tabla donde aparecen las estadísticas de todos los jugadores de todos los equipos de primera, segunda y segunda B. Fuente: Elaboración Propia

Como se puede apreciar, en la primera columna aparece el código del equipo, repitiéndose para cada jugador de dicho equipo. Cuando finalizan los jugadores de un equipo, aparecen seguidamente los del equipo siguiente, en orden creciente del código asignado por la página web. Las siguientes dos columnas son el dorsal y el nombre (deportivo) del jugador. A partir de ahí, se muestran la posición del jugador y todas las estadísticas que ya se indicaron en el apartado anterior. Esto ya es lo que se buscaba.

Como punto final a este subapartado, se debe explicar cómo asigna *BD Fútbol* los códigos de cada uno de los equipos. La web establece los códigos por países o, mejor dicho, por ligas. Los códigos que van del 1 al 1000, son los relativos a los equipos españoles, tanto de primera como de segunda y segunda B.

Investigando detalladamente la estructura de la página web (algo necesario para llevar la implementación de la herramienta a buen puerto), se ha podido comprobar que, además de los códigos de los equipos españoles, *BD Fútbol* asigna el rango 1001-2000 a equipos alemanes, 2001-3000 a equipos ingleses, 3001-4000 a equipos italianos, 4001-5000 a equipos franceses, 5001-6000 a equipos portugueses, 6001-6999 a equipos holandeses y del 10001 en adelante para equipos de otros países y otros continentes.

Sabiendo esto, ya se puede pasar a la siguiente fase de la implementación de la herramienta de análisis en la gestión deportiva.

5.2.2. Desarrollo y tratamiento de la tabla de estadísticas de los jugadores

Una vez se han cargado las tablas de todos los equipos de primera, segunda y segunda división B y de las primeras divisiones de Inglaterra, Alemania, Francia, Italia, Portugal y Holanda, crearemos otro archivo de *Power BI* en el que se desarrollará la primera parte de la herramienta, que consiste en el análisis del



rendimiento de los jugadores del Real Valladolid, siguiendo los criterios que se especificaron en el apartado correspondiente.

En este archivo, sólo se cargará la tabla con las estadísticas de los jugadores del Real Valladolid. En esta parte no hay ninguna complicación para obtener la tabla, pues basta con introducir la dirección URL específica del Real Valladolid. Una vez hecho, yendo a la pestaña de “Datos” podemos ver la tabla. Tiene el siguiente aspecto:

#	Column1	2	Apodo	Pos	PJ	PT	PC	PS	PX	C	Min	TA	TR	G	Minutos/G	GP	GPP	GE	PG	PE	PP
1			Masip	1	35	35	35	0	0	3	3150	2	0	0	0	0	0	39	8	15	12
13			Caro	1	3	3	2	0	1	8	245	0	0	0	0	0	0	3	1	0	2
13			Lunin	1	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0
26			Samu Pérez	1	1	0	0	1	0	8	25	0	0	0	0	0	0	1	0	0	1
2			Pedro Porro	2	13	7	6	6	1	10	737	1	0	0	0	0	0	0	1	5	7
17			Moyano	2	26	26	22	0	4	8	2229	4	0	0	0	0	0	0	9	11	6
34			Corral	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6			Raúl Carnero	3	12	12	10	0	2	4	1009	2	0	0	0	0	0	0	4	3	5
22			Nacho	3	23	21	18	2	3	6	1894	2	0	0	0	0	0	0	5	11	7
3			Barba	4	5	5	5	0	0	4	450	3	0	0	0	0	0	0	1	0	4
5			Javi Sánchez	4	8	6	4	2	2	7	501	1	0	0	0	0	0	0	3	1	4
4			Kiko Olivas	4	35	33	32	2	1	2	2941	2	0	2	1470	0	0	0	7	14	14
24			Joaquín	4	20	19	16	1	3	8	1660	9	0	2	830	0	0	0	5	8	7
30			Alende	4	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
27			Salisu	4	31	30	28	1	2	3	2724	4	0	1	2724	0	0	0	7	14	10
18			Antoñito	5	15	11	9	4	2	12	1030	1	0	0	0	0	0	0	3	6	6
8			Matheus Fernandes	6	3	2	1	1	1	4	166	0	0	0	0	0	0	0	1	0	2
3			Ben Arfa	6	5	2	0	3	2	5	159	1	0	0	0	0	0	0	0	3	2
28			Kike Pérez	6	6	2	1	4	1	1	217	1	0	0	0	0	0	0	1	2	3
6			Luismi	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15			Aguado	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21			Michel	6	29	27	13	2	14	4	2186	6	0	1	2186	1	0	0	7	12	10
8			Anuar	6	11	1	0	10	1	1	183	1	0	0	0	0	0	0	2	5	4
23			Waldo	6	21	9	2	12	7	8	878	6	0	0	0	0	0	0	6	7	8
14			Alcaraz	6	27	20	16	7	4	3	1927	10	0	2	963	0	0	0	7	10	10
11			Hervias	6	25	9	1	16	8	8	1015	1	0	1	1015	0	0	0	5	11	9
19			Toni Villa	6	23	18	2	5	16	12	1357	4	0	0	0	0	0	0	6	10	7
16			Fede San Emeterio	6	25	20	14	5	6	7	1753	4	0	0	0	0	0	0	6	10	9
33			El Hacen	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Tabla 3. Vista de la tabla de estadísticas de los jugadores del Real Valladolid en Power BI. Fuente: Elaboración Propia

No aparecen todos los jugadores del equipo porque la tabla es más grande que lo que se puede ver de una vez. Para ver al resto de jugadores, debemos mover la barra de desplazamiento, situada a la derecha de la tabla. Las columnas, afortunadamente, sí que se pueden visualizar todas al mismo tiempo.

En el apartado anterior, se citaron los indicadores y criterios que se pueden utilizar para evaluar el rendimiento de los jugadores del Real Valladolid. Ahora es el momento de implantarlos en *Power BI*.

Previamente, se crearán en *Excel* dos tablas, que se agregarán al archivo de *Power BI*. La primera de ellas es una leyenda de los términos que aparecen como encabezados de la tabla de las estadísticas de los jugadores. En la página web de *BD Fútbol* se puede ver qué es lo que significa cada uno de esos términos. Una vez creada la tabla y cargada al archivo, el resultado visual es el siguiente:



Siglas	Significado
#	Dorsal
Apodo	Nombre
Pos	Posición
PJ	Partidos Jugados
PT	Partidos Titular
PC	Partidos Completos
PS	Partidos Suplente
PX	Partidos Sustituido
C	Partidos Convocado sin jugar
Min	Minutos jugados
TA	Tarjetas Amarillas
TR	Tarjetas Rojas
G	Goles
Minutos/G	Minutos para anotar un gol
GP	Goles de penal
GPP	Goles en propia puerta
GE	Goles encajados
PG	Partidos Ganados
PE	Partidos Empatados
PP	Partidos Perdidos

Tabla 4. Leyenda con los términos de la tabla de estadísticas de los jugadores. Fuente: Elaboración Propia

Tal y como se puede ver en la tabla 3, la posición de los jugadores viene indicada con un número. El usuario final que deba manipular la herramienta de análisis en la gestión deportiva no tiene porqué conocer el significado de esos números, por lo que se debe incluir otra leyenda en la que se indique cuál es la posición que se corresponde con cada número. Esa es la segunda tabla que se debe crear y agregar al archivo de *Power BI*.

En las tablas que se pueden observar en la web de *BD Fútbol*, la posición de los jugadores no viene representada con un número, sino que se indica con una imagen en la que aparece sombreada una parte del terreno de juego. Observando dichas imágenes y conociendo la posición de la mayoría de los jugadores de La Liga, se ha obtenido la leyenda de las posiciones de los futbolistas.

Número	Posición
1	Portero
2	Lateral Derecho
3	Lateral Izquierdo
4	Central
5	Defensa (Varias posiciones)
6	Centrocampista
7	
8	
9	Delantero Centro
10	Delantero (Varias Posiciones)

Tabla 5. Leyenda de las posiciones de los jugadores. Fuente: Elaboración Propia

Como se puede apreciar, las posiciones 7 y 8 aparecen vacías. Esto se debe a que no hay ningún jugador que tenga ese número en el apartado de posición, desconociéndose el motivo de la no existencia de estos números. Tal vez se deba a que *BD Fútbol* antes usase otra nomenclatura y después decidiera agrupar algunas posiciones o puede que sean posiciones más propias de esquemas tácticos que en la actualidad no se emplean en ningún equipo, pero no se conoce el motivo. De todos modos, no es importante para el desarrollo de la herramienta.



Una vez incluidas estas dos tablas, ya podemos centrarnos en desarrollar la implementación de los indicadores que permitan conocer el rendimiento de los futbolistas del Real Valladolid.

El primer paso por completar consiste en agregar a la tabla de estadísticas de los jugadores una serie de columnas que ayuden a evaluarlos.

La primera de esas columnas será específica para los porteros. En la tabla de estadísticas que *BD Fútbol* ofrece aparece por un lado una columna donde indican los minutos que cada futbolista ha disputado a lo largo de la pasada liga y, por otro lado, viene otra columna donde se muestran los goles encajados por cada jugador. Evidentemente, en todos los jugadores, excepto los porteros, la cifra mostrada es cero. La columna que se creará es combinación de esas dos, pues nos interesa obtener la ratio que mide cada cuántos minutos encaja un gol el portero. En otras palabras, en esa columna debe aparecer el resultado de dividir los minutos jugados entre los goles encajados.

Pero se debe tener cuidado con un aspecto. Si los goles encajados son cero, quedaría una división entre cero, algo que no está permitido en matemáticas. Por lo tanto, es necesario indicar a *Power BI* que sólo debe realizar el cálculo para aquellos jugadores cuya columna de goles encajados tenga algún valor mayor o igual que uno. En caso contrario, debe dejar vacía la casilla correspondiente. Con todas estas indicaciones, la columna se ha creado de la siguiente forma:

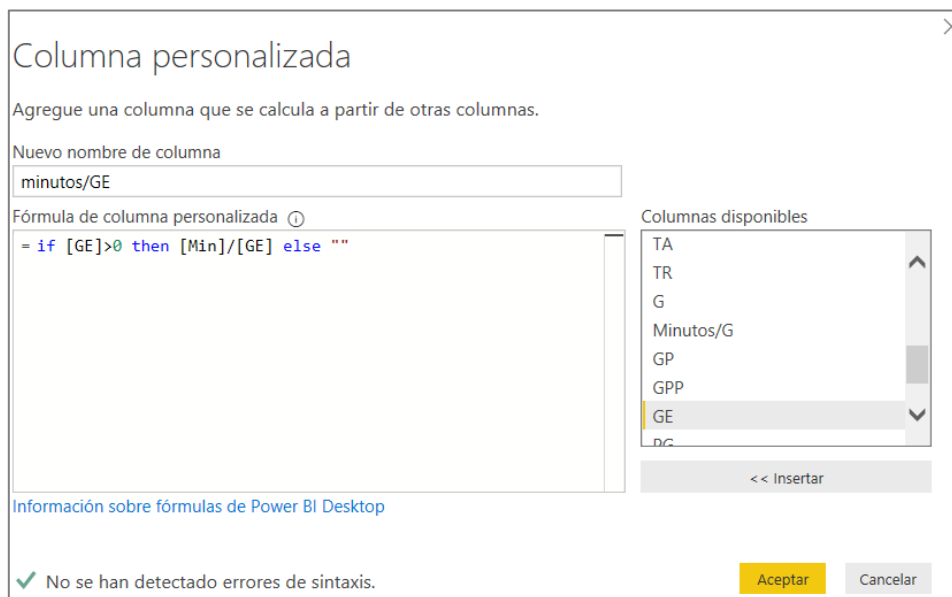


Ilustración 14. Creación de la columna "Minutos/GE". Fuente: Elaboración Propia

La segunda y la tercera columna que se van a insertar siguen un criterio similar. *BD Fútbol* ofrece la columna "Minutos/G" donde calcula los minutos que necesita cada jugador para anotar un gol en promedio. El mismo fin presenta la columna "Minutos/GE" que se acaba de insertar; y también pretenden obtener una ratio equivalente las dos columnas que se van a añadir a continuación.



Estas columnas serán denominadas como “Minutos/TA” y “Minutos/TR”, respectivamente. La primera de ellas calcula cada cuántos minutos recibe un jugador una tarjeta amarilla en promedio, dividiendo para cada futbolista el valor de la columna de los minutos jugados entre el valor de la columna de las amonestaciones recibidas. La segunda calcula cada cuántos minutos recibe un jugador una tarjeta roja en promedio, dividiendo, del mismo modo, para cada futbolista el valor de la columna de los minutos jugados entre el valor de la columna de las expulsiones con las que ha sido sancionado.

Al igual que se expuso antes, es necesario tener en consideración algunas restricciones que se deben indicar al programa, para evitar obtener valores indeseados. Si un jugador no ha disputado ningún minuto, lógicamente tampoco habrá recibido ninguna tarjeta (ni amarilla ni roja). Por consiguiente, si el programa realiza ese cálculo obtendría la indeterminación cero entre cero. Para sortear este problema hay que indicar a *Power BI* que, si los minutos de un determinado jugador son cero, debe dejar en blanco sus casillas de las columnas que se van a agregar.

Seguidamente, nos tenemos que centrar en los jugadores que sí que han disputado algún minuto. Entre estos futbolistas existen dos posibilidades. La primera de ellas es que sí que hayan recibido alguna amarilla y/o alguna roja; y en esos casos es en los que se quiere averiguar cada cuántos minutos ha sido sancionado el jugador. La segunda posibilidad es que el jugador haya disputado minutos pero que no haya sido amonestado ni expulsado; presentándose el problema de que el resultado sería una indeterminación del tipo $\frac{x}{0}$. Para solucionarlo, se debe indicar que sólo se debe ejecutar la operación en el caso de aquellos jugadores cuyo valor de tarjetas amarillas (y rojas en el caso correspondiente) recibidas sea mayor o igual que uno. Con todo lo expuesto, la programación para agregar la columna que evalúa cada cuántos minutos es amonestado cada futbolista queda así:

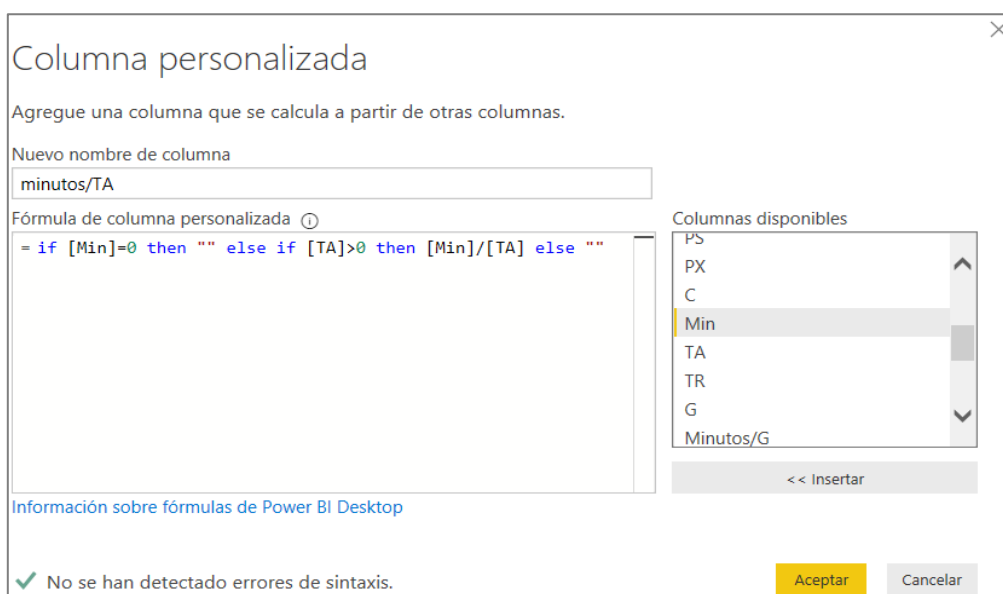


Ilustración 15. Creación de la columna "Minutos/TA". Fuente: Elaboración Propia



Y, de una forma idéntica, la programación escrita para insertar la columna que evalúa cada cuántos minutos es expulsado cada jugador del equipo es la que se muestra en la siguiente ilustración:

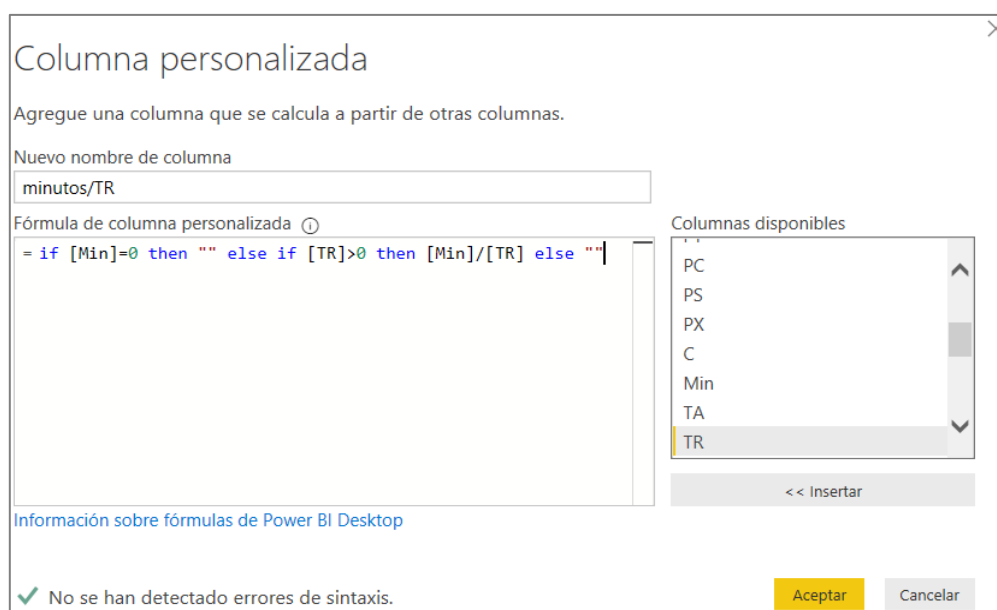


Ilustración 16. Creación de la columna "Minutos/TR". Fuente: Elaboración Propia

Como ya se comentó en puntos anteriores, en el caso particular de los jugadores del Real Valladolid, nadie recibió ninguna tarjeta roja a lo largo de la temporada 2019/2020. Por lo tanto, esa columna no contendrá ningún valor. Sin embargo, ésta se ha incluido porque será necesaria cuando haya que evaluar a los jugadores del resto de los equipos y, a mayores, sería necesaria en el caso de que la herramienta se fuera a utilizar para futuras temporadas; idea que se ha estado valorando durante todo el proceso de confección de la herramienta de análisis en la gestión deportiva.

Una vez han sido creadas las tres columnas, es conveniente especificar el tipo de datos que contienen, para evitar posibles complicaciones que puedan surgir después, por ejemplo, a la hora de agregar algunas visualizaciones.

Cuando se importan tablas a *Power BI*, sea desde la fuente que sea, el programa establece por defecto el tipo de datos que tiene cada columna de esa tabla, siempre que los valores que tengan sean más o menos homogéneos (si es texto, que todas las filas de esa columna contengan texto, si son números enteros, que todas las filas tengan números enteros...). Sin embargo, cuando se crean manualmente las columnas, no aparece ningún tipo, por lo que se debe indicar de forma manual. Los tipos de datos que puede interpretar *Power BI* son los siguientes:

- Número decimal.
- Número decimal fijo.
- Número entero.
- Porcentaje.



- Fecha y hora.
- Fecha.
- Hora.
- Fecha, hora y zona horaria.
- Duración.
- Texto.
- Verdadero o falso.
- Binario.

En el caso de las tres columnas que se han creado, al tratarse de la división de dos números enteros cualesquiera, se indicará que todas ellas contienen datos del tipo “número decimal”. Una vez hecho, el resultado visual es el siguiente:

Apellido	Pos	PJ	PT	PC	PS	PX	C	Min	TA	TR	G	Minutos/G	GP	GPP	GE	PG	PE	PP	minutos/GE	minutos/TA	minutos/TR
Masip	1	35	35	35	0	0	3	3150	2	0	0	0	0	0	35	8	15	12	80,7692307692308		1575
Caro	1	3	3	2	0	1	8	245	0	0	0	0	0	0	3	1	0	2	81,6666666666667		
Lunin	1	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0			
Samu Pérez	1	1	0	0	1	0	8	25	0	0	0	0	0	0	1	0	0	1	25		
Pedro Porro	2	13	7	6	6	1	10	737	1	0	0	0	0	0	0	1	5	7		737	
Moyano	2	26	26	22	0	4	8	2229	4	0	0	0	0	0	0	9	11	6		557,25	
Corral	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Raúl Carrero	3	12	12	10	0	2	4	1009	2	0	0	0	0	0	0	4	3	5		504,5	
Nacho	3	23	21	18	2	3	6	1894	2	0	0	0	0	0	0	5	11	7		947	
Barba	4	5	5	5	0	0	4	450	3	0	0	0	0	0	0	1	0	4		150	
Javi Sánchez	4	8	6	4	2	2	7	501	1	0	0	0	0	0	0	3	1	4		501	
Kiko Olivás	4	35	33	32	2	1	2	2941	2	2	1470	1470	0	0	0	7	14	14		1470,5	
Joaquín	4	20	19	16	1	3	8	1660	9	2	830	830	0	0	0	5	8	7		184,444444444444	
Alende	4	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0			
Salisu	4	31	30	28	1	2	3	2724	4	0	1	2724	0	0	0	7	14	10		681	
Antofrito	5	15	11	9	4	2	12	1030	1	0	0	0	0	0	0	3	6	6		1030	
Matheus Fernandes	6	3	2	1	1	1	4	166	0	0	0	0	0	0	0	1	0	2			
Ben Arfa	6	5	2	0	3	2	5	159	1	0	0	0	0	0	0	0	3	2		159	
Kike Pérez	6	6	2	1	4	1	1	217	1	0	0	0	0	0	0	1	2	3		217	
Luismi	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Aguado	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Michel	6	29	27	13	2	14	4	2186	6	1	2186	2186	1	0	0	7	12	10		364,333333333333	
Anuar	6	11	1	0	10	1	1	183	1	0	0	0	0	0	0	2	5	4		183	
Waldo	6	21	9	2	12	7	8	878	6	0	0	0	0	0	0	6	7	8		146,333333333333	
Alcaraz	6	27	20	16	7	4	3	1927	10	0	2	963	0	0	0	7	10	10		192,7	
Hervías	6	25	9	1	16	8	8	1015	1	0	1	1015	0	0	0	5	11	9		1015	
Toni Villa	6	23	18	2	5	16	12	1357	4	0	0	0	0	0	0	6	10	7		339,25	
Fede San Emeterio	6	25	20	14	5	6	7	1753	4	0	0	0	0	0	0	6	10	9		438,25	
El Morán	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0			

Tabla 6. Tabla de estadísticas de los jugadores del Real Valladolid en la temporada 2019/2020, una vez se han agregado las columnas "Minutos/GE", "Minutos/TA" y "Minutos/TR". Fuente: Elaboración Propia

Una vez hecho esto, ya está la tabla casi completa y disponible para poder crear el informe sobre el rendimiento de los jugadores del Real Valladolid e insertar en él todas las visualizaciones que se consideren oportunas. Sólo queda una última columna por agregar, que posiblemente sea la más importante de cara a la ayuda a los responsables del área de planificación deportiva del club para llevar a cabo la toma de decisiones. Esa columna no es otra que la correspondiente a la valoración de los jugadores.

Si cogemos un determinado volumen de datos de una base de datos, tratamos esos datos y sacamos unas gráficas, tablas... podremos justificar que hemos hecho análisis de datos. Sin embargo, este trabajo de fin de grado también trata de la inteligencia de negocio o inteligencia empresarial. Por lo tanto, es necesario justificar que también se ha ofrecido al cliente, el Real Valladolid, una herramienta donde se desarrolla inteligencia de negocio.



En otras palabras, se podría decir que hay que demostrar porqué es mejor acudir a la herramienta desarrollada en este TFG que acudir directamente a la web de *BD Fútbol*. Y la única manera de conseguirlo es ofreciendo algo a mayores de la información proporcionada por la web. La primera información que se ofrece en la herramienta y no en la web es la columna de valoración de los jugadores.

Con esta columna se pretende que, los responsables de decidir qué jugadores deben ser renovados, qué jugadores deben ser vendidos y cuáles deben ser fichados, puedan obtener una primera impresión de cada jugador con sólo ver la valoración de su rendimiento durante la pasada temporada. Lógicamente, no se podrán basar únicamente en este criterio, pues son muchos los factores que entran en juego; pero sí que puede facilitar mucho el trabajo a la hora de hacer un primer cribado de jugadores. Por ejemplo, los responsables podrían eliminar directamente a todos aquellos futbolistas que no superen una determinada valoración y, de este modo, ahorrarse muchas horas de trabajo con un riesgo relativamente bajo de no considerar a un buen futbolista.

Para calcular esta columna se ha procedido de la siguiente forma: Aquellos jugadores que no hayan disputado ningún partido con un determinado equipo tendrán una valoración de 1 (si luego han jugado con algún otro club tendrán la valoración que corresponda a su rendimiento en dicho club). Para aquellos que sí que hayan disputado algún partido, la valoración de su rendimiento se obtendrá como suma del valor obtenido en cada uno de los parámetros que mide *BD Fútbol*, los cuales ya se comentaron al principio del caso práctico. Algunos ejemplos son los siguientes:

- Si el jugador ha participado en más de 31 partidos obtiene 3 puntos, si ha participado en más de 24 partidos obtiene 2 puntos, si ha participado en más de 11 partidos obtiene 1 punto y, en caso contrario, no obtiene ningún punto. Con ello, se pretende premiar su regularidad a lo largo de la temporada.
- Si el futbolista ha sido suplente en más de 14 encuentros no obtiene ningún punto, si ha sido suplente en más de 6 encuentros obtiene 1 punto y, en el caso de haber partido como suplente en 6 o menos ocasiones, obtiene dos puntos. En este caso se busca premiar a los jugadores que son importantes para el entrenador y que casi siempre que están, juegan.
- Los jugadores que anotan un gol cada 600 o más minutos en promedio obtienen 1 punto, pues eso quiere decir que sí que marcan, pero que necesitan mucho tiempo para hacerlo. Los que necesitan entre 400 y 599 minutos para lograr un tanto, obtienen 2 puntos, ya que sigue sin ser un registro espectacular, pero está mejor que lo anterior. Los que necesitan menos de 400 minutos para marcar obtienen 3 puntos, pues son los buenos goleadores. Y, por último, los que no han anotado ningún gol, no suman ningún punto.



Con una lógica similar a la de los ejemplos expuestos, se ha ido estableciendo una ponderación para cada parámetro evaluado en las tablas de estadísticas de los jugadores que *BD Fútbol* ofrece. De ese modo, se ha implementado en *Power BI* la columna de valoración global de los jugadores, como se puede ver en la siguiente ilustración:

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna

Fórmula de columna personalizada ⓘ

```
= if [PJ]=0 then 1 else ((if [PJ]>31 then 3 else if [PJ]>24 then 2 else if [PJ]>11 then 1 else 0) + (if [PT]>29 then 3 else if [PT]>19 then 2 else if [PT]>9 then 1 else 0) + (if [PC]>23 then 3 else if [PC]>15 then 2 else if [PC]>8 then 1 else 0) + (if [PS]>14 then 0 else if [PS]>6 then 1 else 2) + (if [PX]>9 then 0 else if [PX]>3 then 1 else 2) + (if [C]>7 then 0 else if [C]>2 then 1 else 2) + (if [Min]>2999 then 3 else if [Min]>1999 then 2 else if [Min]>999 then 1 else 0) + (if [TR]>3 then 0 else if [TR]>0 then 1 else 2) + (if [#"Minutos/G"]>599 then 1 else if [#"Minutos/G"]>399 then 2 else if [#"Minutos/G"]>0 then 3 else 0) + (if [PG]
```

Columnas disponibles

- #
- 2
- Apodo
- Pos
- PJ
- PT
- PC

<< Insertar

✓ No se han detectado errores de sintaxis. Aceptar Cancelar

Ilustración 17. Columna donde se calcula una valoración global del rendimiento de cada jugador. Fuente: Elaboración Propia

Es importante indicar que no se han valorado todas las columnas, porque eso daría lugar a duplicados. Por ejemplo, si se valorara la columna de goles y la columna de minutos necesarios para anotar un gol, se estaría premiando a aquellos jugadores que marcan mucho dos veces por “lo mismo”, puesto que, cuantos más goles marcas, menos minutos te ha costado hacer cada tanto. Es decir, ambos indicadores son directamente proporcionales. En este caso, se ha decidido valorar la columna de minutos necesarios para anotar un gol, puesto que es un criterio más objetivo.

Y es más objetivo porque, por ejemplo, decir que un futbolista ha anotado 7 goles es dar una información escasa. Es decir, si ha anotado 7 goles en 36 partidos no es una cifra especialmente buena, pero si los ha marcado en tan solo 10 encuentros entonces es un resultado más que favorable. Por el contrario, si se calcula la ratio minutos jugados entre goles anotados, da igual los partidos que haya disputado el jugador, puesto que a todos los futbolistas se les está valorando bajo el mismo criterio.

Lo mismo sucede en el caso de los goles encajados y los minutos necesarios para encajar un gol y en el caso de las tarjetas amarillas recibidas y los minutos necesarios para ser amonestado. En ambos se ha usado el segundo indicador, por la razón expuesta anteriormente.



Sin embargo, se ha hecho una excepción en el caso de las tarjetas rojas recibidas y los minutos necesarios para ser expulsado en promedio. Aquí se ha decidido valorar ambos indicadores por dos razones. En primer lugar, por la importancia que tiene en el juego el hecho de que un equipo se quede con un hombre menos. No se trata de premiar a los que nunca son expulsados, que también, sino de penalizar a aquellos jugadores que son expulsados con relativa frecuencia. El segundo motivo es que, entre los jugadores que no han recibido ninguna tarjeta roja o que han recibido una y los jugadores que más rojas hayan visto no hay, en casi ningún caso, mucha diferencia, pues es muy raro ver a un futbolista que es expulsado más de 3 o 4 veces en una misma temporada. Por lo tanto, si se tuviera en cuenta la ratio de minutos jugados entre rojas recibidas es probable que hubiera poca diferencia entre, por ejemplo, un jugador que ha sido expulsado una vez y otro que lo ha sido tres veces, cuando la diferencia práctica o real es enorme.

Por último, hay que indicar que no se han tenido en cuenta los goles anotados en tiros penales, puesto que no es un criterio justo, ya que, en la mayoría de los equipos es un único jugador el que tira los penales o entre dos o tres jugadores a lo sumo. Por lo tanto, no todos tendrían opción de puntuar. Además, esos goles están incluidos en el indicador de minutos necesarios para anotar un gol, que sí se valora. Por lo que, al igual que se explicaba anteriormente, si se considerara estaríamos cayendo en una duplicidad.

Tampoco se han considerado los goles anotados en propia puerta, por diversos criterios subjetivos. Por ejemplo, que es injusto tanto premiar a un jugador que no se ha metido en propia ningún gol, quizás porque apenas pisa su área, como penalizar a un jugador que sí que se haya metido alguno, porque quizás fue mala suerte, un golpe fortuito o tratando de evitar un gol del equipo adversario que habría entrado de igual forma, pero que en nada tiene porqué empañar su actuación.

Con todo lo expuesto, se puede afirmar que, en este punto, ya está todo listo para crear y desarrollar el informe de la herramienta de análisis en la gestión deportiva. Es decir, ahora es el momento de implementar todos aquellos paneles, gráficos, visualizaciones, etc. con los que trabajará el cliente una vez le sean entregados a la conclusión del proyecto.

5.2.3. Creación del informe sobre el análisis de los futbolistas

Desde la perspectiva del que lleva a cabo el proyecto, es muy probable que esta parte sea la más determinante de cara a obtener la satisfacción y reconocimiento del cliente. En definitiva, es el informe lo que le será entregado.

El cliente, en este caso el Real Valladolid, no ve el trabajo que hay detrás del informe. Las horas dedicadas a estudiar el programa que se va a utilizar (*Power BI* en este supuesto planteado), a analizar y comprender el funcionamiento y la estructura de la base de datos de la que se nutre al programa (*BD Fútbol* en nuestro caso), a planificar y decidir que elementos se utilizarán y cómo se mostrarán y demás cuestiones imprescindibles para desarrollar e implementar la herramienta de



análisis en la gestión deportiva son horas que nadie valorará si el resultado final no es efectivo y atractivo. Sí, atractivo. Cada vez más, las tendencias exigen a los desarrolladores crear aplicaciones y herramientas que sean visualmente agradables y cómodas, para acelerar y facilitar al cliente el proceso de toma de decisiones.

A continuación, se explicará cómo se ha creado el informe para analizar el rendimiento de los jugadores del Real Valladolid, indicando, panel a panel, qué se ha puesto y el motivo por el cuál se ha incluido cada visualización. En otras palabras, se comentará qué objetivos se pretenden conseguir con cada una de las páginas del informe. Posteriormente se comentarán las diferencias que hay en el segundo informe (el de análisis y comparación de los jugadores del resto de equipos con los jugadores del Real Valladolid), que en principio es muy similar.

La primera página es la portada. En ella, aparece el título del informe, *Análisis del Rendimiento de los Jugadores del Real Valladolid*, el nombre del autor y unas imágenes insertadas que cumplen la función de botones, ya que permiten al usuario avanzar a cada una de las páginas del informe. Todo ello con los colores típicos del club: el blanco y el violeta. De fondo, se ha insertado una imagen de color violeta oscuro degradado hacia el centro, donde se puede ver el escudo del Real Valladolid en tonos monocromáticos. La portada es la siguiente:



Ilustración 18. Portada del informe del análisis del rendimiento de los jugadores del Real Valladolid. Fuente: Elaboración Propia

Para que pulsando en cada una de las imágenes que hacen de botones *Power BI* nos lleve a la página correspondiente, se debe proceder de la siguiente manera: En primer lugar, una vez se ha desarrollado el informe entero, se tienen que crear los denominados marcadores. Para ello, se siguen los siguientes pasos:

Vista → Panel de Marcadores → Agregar



Nos aparecerá un nuevo marcador, en el cual tenemos que indicar el nombre que queremos que tenga. Una vez hecho esto, sólo queda acudir a la página del informe a la que queremos que nos lleve, pulsar sobre los tres puntos que aparecen a la derecha del nombre del marcador y dar a “Actualizar”, para que el programa registre que ese marcador debe llevar a esa página.

Hecho esto, ya está creado correctamente el marcador. El segundo y último paso consiste en asociar el marcador creado con la imagen insertada. Para ello, pinchamos sobre la imagen y se desplegará a la derecha de la pantalla el menú, llamado “Formato de Imagen”. Ahí se verá que, por defecto, “Acción” aparece deshabilitado. Debemos activarlo y desplegar el submenú correspondiente. Debemos elegir en primer lugar el tipo de acción que queremos, que puede ser:

- “Atrás”: Para que al pulsar nos lleve a la pestaña anterior.
- “Marcador”: Para indicar el marcador que se desee.
- “Preguntas y respuestas”.
- “URL web”: Donde se puede insertar un enlace web al que el programa nos conduciría en caso de pulsar sobre la imagen correspondiente.

En el supuesto planteado se elegirá la opción “Marcador”. El programa solicitará que se especifique cuál es el marcador que se quiere asociar a esa imagen de todos los que haya creados, se indica y ya se habría terminado.

La segunda página del informe es el comparador de jugadores de campo. Al igual que en todas las páginas siguientes, en la parte superior se ha puesto un encabezado consistente en una franja violeta partida en su lado derecho por el escudo del Real Valladolid. A parte de dar un toque visual y atractivo al informe, este encabezado también cumple las funciones de botón pues, siguiendo el mismo procedimiento que se ha indicado anteriormente, se ha establecido la portada como marcador de esta imagen. Es decir, si pulsamos sobre ella, *Power BI* nos conducirá a la portada del informe.

El resto de la página se ha estructurado en tres partes. Las dos partes laterales corresponden con la “ficha” de cada uno de los jugadores que se van a comparar, mientras que la parte central contiene dos velocímetros para cada jugador.

La ficha de cada jugador contiene, en grande y en la parte superior, la foto del jugador y la valoración global de éste. Abajo aparecen seis indicadores del rendimiento del jugador durante la pasada temporada 2019/2020. Son los siguientes:

- Partidos jugados.
- Partidos en los que parte como titular.
- Minutos jugados a lo largo de la temporada.
- Partidos ganados.



- Goles anotados.
- Tarjetas amarillas recibidas.

En la parte central aparecen dos secciones. En la superior se evalúa a cada jugador en función de los minutos que necesita, en promedio, para anotar un gol. Y en la inferior se evalúa a cada jugador en función de los minutos que tarda, en promedio, en recibir una tarjeta amarilla. En ambos casos se han agregado unos velocímetros, ya que, además de indicar numéricamente los minutos en cada medida, también tienen un fuerte impacto visual, ya que lo primero que aprecia el usuario cuando lo mira es la posición en la que se sitúa la flecha. Si está en la parte verde quiere decir que el resultado es bueno, si está en la parte amarilla significa que está en una zona intermedia y, si por el contrario la flecha se sitúa en la zona roja, es que el resultado es malo.

En el caso de los minutos necesarios para anotar un gol, sería bueno que tardara el mínimo tiempo posible en anotar un tanto y sería peor cuantos más minutos necesitara el jugador para marcar. Sucede lo contrario en el caso de los minutos que tarda en ser amonestado cada futbolista. Cuantos menos minutos requiera para ver una amarilla peor.

En la siguiente ilustración se puede ver el descrito comparador de jugadores de campo, donde, a modo de ejemplo, aparece la comparativa entre Sergi Guardiola y Enes Ünal, los dos principales delanteros del equipo durante la pasada campaña.



Ilustración 19. Comparador de jugadores de campo. Fuente: Elaboración Propia

El texto “Comparativa del Rendimiento de los Jugadores” ha sido implementado mediante un cuadro de texto, cuyo funcionamiento es igual al que pueden tener los cuadros de texto que se pueden encontrar en Excel. Éstos permiten editar el tipo de letra, el tamaño, el color, la posición dentro del cuadro, si contiene



un enlace web, etc. Del mismo modo se han incluido los textos “*Minutos Necesarios para Anotar un Gol*” y “*Minutos Necesarios para Recibir una Amarilla*”.

Los seis indicadores que aparecen en la ficha de cada jugador, así como las valoraciones globales de cada uno han sido implementados mediante la visualización “*Tarjeta*”. A esta visualización se la debe pasar un campo; por ejemplo, en el caso de los partidos jugados, se le pasará el campo “*PJ*” de la tabla de estadísticas de los jugadores. En el menú de configuración de cada tarjeta se pueden editar muchísimos aspectos. Entre otras cosas, se puede establecer un título para la tarjeta (por ejemplo, “*Partidos Como Titular*”, en el caso de la etiqueta con el campo “*PT*”); el color de fondo, si queremos que lo tenga, y el tanto por ciento de transparencia, el tamaño y la posición de la tarjeta; el tamaño, la posición, el tipo de letra y el color del texto y de los números...

La imagen con el rostro de los jugadores se ha agregado a través de la visualización “*Simple Image*”. Por su parte, los velocímetros corresponden a la visualización “*Tachometer*”, la cual no viene por defecto instalada en el programa, si no que es necesario descargarla en lo que *Power BI* denomina como el “*Marketplace*”.

A esta última visualización es necesario pasarle varios valores. En primer lugar, se le debe indicar la medida que se va a mostrar tanto en la posición de la flecha como numéricamente en la parte inferior del velocímetro. En el caso de los minutos necesarios para anotar un gol se le pasará la columna “*Minutos/G*” de la tabla de estadísticas de los jugadores. Por su parte, en el caso de los minutos que tarda en recibir una tarjeta amarilla en promedio, se le indicará que la columna correspondiente es “*Minutos/TA*”. A mayores, es necesario pasarle, al menos, cuatro valores más. Son los siguientes:

- “*Start Value*”: Es el valor en el que empieza el velocímetro.
- “*End Value*”: Es el valor en el que termina el velocímetro.
- “*Range2 Start Value*”: Es el valor en el que termina el primer tramo (el de color verde en el caso de los minutos necesarios para anotar un gol y el de color rojo en el caso de los minutos que tarda en recibir una amarilla) y comienza el segundo (el de color amarillo).
- “*Range3 Start Value*”: Es el valor en el que termina el segundo tramo y comienza el tercero (el de color rojo en el caso de los minutos necesarios para anotar un gol y el de color verde en el caso de los minutos que tarda un jugador en recibir una amonestación).

También se le podría pasar el parámetro “*Target Value*” que sería el objetivo que se quiere alcanzar en cada indicador. Sin embargo, se ha decidido no establecer ningún objetivo puesto que habría que indicar uno para los delanteros, otro para los centrocampistas y otro para los defensas; y esta visualización sólo admite un valor. Todos estos valores que se le deben pasar se registran manualmente como medidas,

pulsando sobre “Nueva Medida”, situada en la parte de “Cálculos” de la pestaña de “Inicio”.

Para poder comparar jugadores es necesario poder filtrarlos y seleccionar los que queramos en cada circunstancia. Para ello, se ha empleado el elemento “Segmentación de Datos”, situado también en la parte de visualizaciones. A este elemento hay que pasarle únicamente un campo que, en este caso, no es otro que el del apodo de los jugadores.

En el menú de configuración y edición se pueden modificar muchas cosas, como por ejemplo la presencia o no de un fondo, así como su color y tanto por ciento de transparencia; establecer un título, indicando el tipo, tamaño, posición y color de la letra; establecer los controles de selección, como permitir una selección múltiple o, por el contrario, obligar a fijar una única selección, etc. Pero lo más importante es que, al introducir el nombre de un jugador, sólo se filtren los velocímetros, las tarjetas y la foto de un lado de la pantalla y no todos (porque en el otro lado se querrá poner, lógicamente, los datos de otro jugador). Para lograr esto, se pulsa sobre el filtro correspondiente y se procede de la siguiente manera:

Herramientas de objetos visuales → Formato → Editar Interacciones

Una vez seguidos esos pasos aparecerán, sobre cada elemento de la página, dos iconos. El de la izquierda, que es el que viene seleccionado por defecto, sirve para permitir la interacción del filtro que estamos editando con ese elemento. Por el contrario, el icono de la derecha se debe pulsar cuando queramos que el filtro no afecte a ese elemento. Es decir, cuando queremos que una visualización no cambie al modificar el jugador de ese filtro. En la siguiente ilustración, se ve el resultado de haber hecho esto en el filtro de la izquierda:

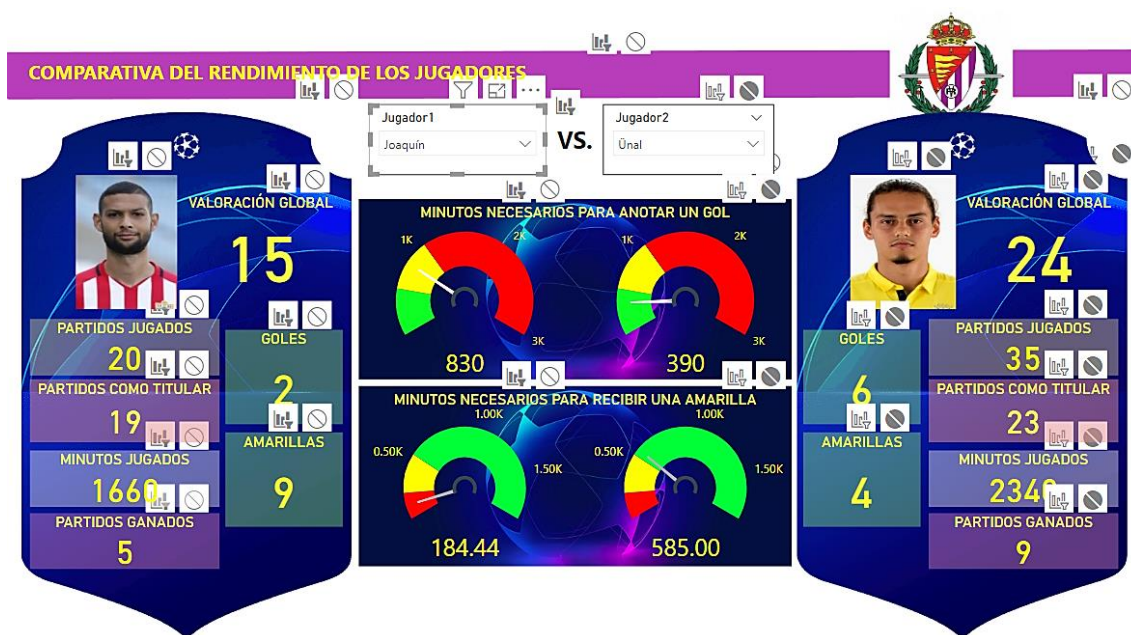


Ilustración 20. Resultado de editar las interacciones del filtro de la izquierda. Fuente: Elaboración Propia



Una vez llevada a cabo esta edición en ambos filtros de jugadores, sólo queda un paso para dar por finalizado este comparador de jugadores.

Tal y como se indicó al empezar a explicar esta página, éste es un comparador para jugadores de campo. Por lo tanto, no es apto para la evaluación de los porteros. Por consiguiente, se debe especificar a *Power BI* que, en esta página, sólo se deben poder seleccionar defensas (tanto centrales como laterales izquierdos y derechos), centrocampistas y delanteros. Para lograr esto, se debe crear un filtro general para toda la página, indicando que están habilitadas todas las posiciones excepto la de portero.

Esto se puede hacer en la pestaña de “*Filtros*”, situada entre la página del informe y la pestaña de visualizaciones. Se arrastra el campo correspondiente, en el supuesto planteado es la columna “*Pos*” de la tabla de estadísticas de los jugadores, y se selecciona un filtrado básico, en el que se marcan todas las posiciones excepto la 1, que es la relativa a la demarcación de portero.

La tercera página del informe es el comparador de porteros. Puesto que, como se especificará a continuación, es muy similar al comparador de jugadores de campo, para crearlo se ha duplicado directamente la página anterior y se ha procedido a modificar sólo aquello que varía de uno a otro.

Lo primero que se ha modificado es el cuadro de texto del encabezado, donde se ha sustituido “*Comparativa del Rendimiento de los Jugadores*” por “*Comparativa del Rendimiento de los Porteros*”, y el cuadro de texto donde ponía “*Minutos Necesarios para Anotar un Gol*”, que se ha cambiado por “*Minutos Necesarios para Encajar un Gol*”.

Lo segundo que se ha hecho es cambiar el filtro de la página. Si en el caso anterior se permitían todas las posiciones excepto la de portero, ahora debe suceder exactamente lo contrario. Es decir, se tiene que indicar exclusivamente la posición 1, relativa a la demarcación de portero.

La tercera, y última, modificación que hay que ejecutar es la consistente en cambiar los parámetros de los velocímetros que medían, para cada jugador, los minutos que necesitaba en promedio para anotar un gol. También habrá que cambiar la etiqueta de goles anotados, por la correspondiente de goles encajados.

En su lugar, se medirán los minutos que ha tardado cada portero en encajar un gol en promedio. Por lo que, en este caso, al velocímetro se le pasará la columna “*Minutos/GE*”. Del mismo modo, se eliminarán las medidas correspondientes a los goles marcados y se insertarán las medidas creadas para definir los rangos relativos a los minutos necesarios para encajar un gol.

Una vez hecho esto, este comparador ya estaría listo, siendo su aspecto el que se puede apreciar en la siguiente imagen, donde se compara a los dos porteros del Real Valladolid que más partidos disputaron durante la pasada liga:



Ilustración 21. Comparador de Porteros. Fuente: Elaboración Propia

La cuarta y la quinta página son las correspondientes a las matrices estratégicas, aunque también podrían haberse denominado matrices comparativas. Aún sin ser tan visuales o llamativas como los comparadores, estos gráficos tienen una notable importancia de cara a la ayuda en la toma de decisiones. De un vistazo, observando la posición de cada jugador en el gráfico, los responsables pueden ver cuáles son los puntos débiles de un jugador, en qué aspectos destaca sobre los demás futbolistas y en qué parámetros está en la media del equipo.

En la cuarta página aparecen dos matrices. La primera de ellas evalúa la relación entre los minutos jugados y la valoración global obtenida. El tamaño del círculo de cada jugador es proporcional a las victorias en las que participó éste. Por su parte, la segunda mide la relación entre los minutos jugados y las tarjetas amarillas recibidas. Y, al igual que en la anterior, el tamaño de los círculos de los jugadores es función del número de victorias en las que estuvo presente cada uno de ellos.

La visualización que se ha empleado para representar estas matrices es “*Quadrant Chart by MAQ Software*”. Del mismo modo que sucedía con los velocímetros, esta visualización se ha descargado del “*Marketplace*” de *Power BI*, pues no aparecía por defecto en la versión de escritorio del programa. Esta visualización requiere que le pasen varios elementos. Son los siguientes:

- “X Axis”: Aquí debe ir el elemento que irá en el eje X. En ambas gráficas son los minutos jugados por cada futbolista.
- “Y Axis”: En este caso se le pasará el elemento que ocupará el eje Y. En la matriz de la izquierda es la valoración global obtenida y en la matriz de la derecha es el número de tarjetas amarillas recibidas.



- “Radial Axis”: En este campo se debe introducir el parámetro en función del cual se calculará el tamaño de cada círculo. En ambas matrices se calculará en función de los partidos ganados.
- “Legend Axis”: Este campo es el correspondiente a los elementos que serán evaluados. Es decir, los jugadores. Por lo tanto, se le pasará la columna “Apodo” a ambas matrices.

Si se acude al apartado de configuración, se pueden modificar muchas cosas. Por ejemplo, se puede indicar el texto que queremos que aparezca en cada cuadrante, se puede modificar el punto de comienzo y final de cada eje, así como su intervalo, se pueden cambiar los colores de los círculos, el título, el color de fondo y su tanto por ciento de transparencia, etc. Con todo ello, el resultado es el que se muestra a continuación:



Ilustración 22. Matrices Estratégicas I. Fuente: Elaboración Propia

Como se puede apreciar, también aparecen dos filtros en la parte superior de la página. El primero de ellos es para filtrar por posición y el segundo es para filtrar por nombre del jugador. Si bien es cierto que no es necesario filtrar en el caso del informe acerca de los jugadores del Real Valladolid (debido a que hay poco más de veinte jugadores), sí que será un elemento a tener en cuenta cuando se cargue la macro-tabla con los jugadores de todos los demás equipos, ya que, la presencia de miles de futbolistas hará inviable representarlos a todos en una misma gráfica.

En la quinta página se podrán encontrar otras dos gráficas (mejor que matrices en este caso). La gráfica de la izquierda trata la relación entre los partidos jugados y los partidos como titular. Al mismo tiempo, el tamaño de los círculos de cada jugador es función de las tarjetas amarillas recibidas. La segunda gráfica mide la relación entre los partidos jugados y los partidos ganados, siendo en este caso los goles el parámetro que determina el tamaño del círculo de cada futbolista del Real Valladolid.



En esta página, la visualización implementada es, en ambos casos, un gráfico de dispersión. A diferencia de las matrices anteriores, este tipo de visualización sí que viene instalada por defecto en la aplicación de escritorio de *Power BI*. De una forma parecida a las matrices, los elementos que necesitan ser indicados a los gráficos de dispersión son los siguientes:

- “*Leyenda*”: Son los elementos que aparecerán en el gráfico. En otras palabras, es lo que se evalúa. En ambas gráficas serán los jugadores, por lo que se le pasará el campo “*Apodo*”.
- “*Eje X*”: Aquí se debe indicar el elemento que ocupará el eje X. En ambas gráficas son los partidos jugados.
- “*Eje Y*”: Aquí se debe indicar el campo que estará representado en el eje Y. En la gráfica de la izquierda serán los partidos como titular de cada jugador y, en la gráfica de la derecha, serán los partidos ganados por cada futbolista.
- “*Tamaño*”: En este caso se deberá especificar el campo en función del cual queremos que determine el tamaño del círculo relativo a cada uno de los jugadores del equipo.

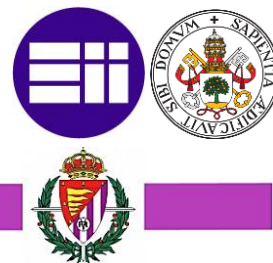
Los gráficos de dispersión también permiten que se les pasen otros parámetros, que son “*Detalles*” y “*Eje de Reproducción*”. Sin embargo, no serán necesarios en el supuesto planteado.

Si se acude al menú de edición y configuración, son muchas las cosas que se pueden poner, quitar y/o modificar. Se puede establecer el volumen máximo de datos admitidos, la presencia o no de leyenda, así como su posición y tamaño, el tipo de escala de cada uno de los ejes, los colores y formas de los datos, el área de trazado (del cual se hablará a continuación), la presencia o no de etiquetas, el título, pudiendo modificar su contenido, tamaño, tipo de letra, color y posición, y un larguísimo etcétera.

En lo que respecta a la determinación del área de trazado, esta visualización permite que se incluya como fondo una imagen, pudiendo determinar el grado de transparencia de ésta. Con el objetivo de obtener un resultado visualmente más atractivo y agradable, se ha decidido incluir una imagen del escudo del Real Valladolid, estableciendo un 80 % de transparencia para que no entorpezca la observación y el análisis de los datos, que es lo fundamental. También conviene mencionar que se puede indicar el tipo de ajuste de imagen que se desee.

Por último, se menciona que en esta página se han incluido los mismos filtros que en la anterior y, al igual que se explicaba antes, su utilidad no estará en esta parte de la herramienta, si no a la hora de evaluar el rendimiento de los jugadores de todos los demás clubes.

Con todo ello, el resultado es el que aparece en la siguiente ilustración:



MATRICES ESTRATÉGICAS II

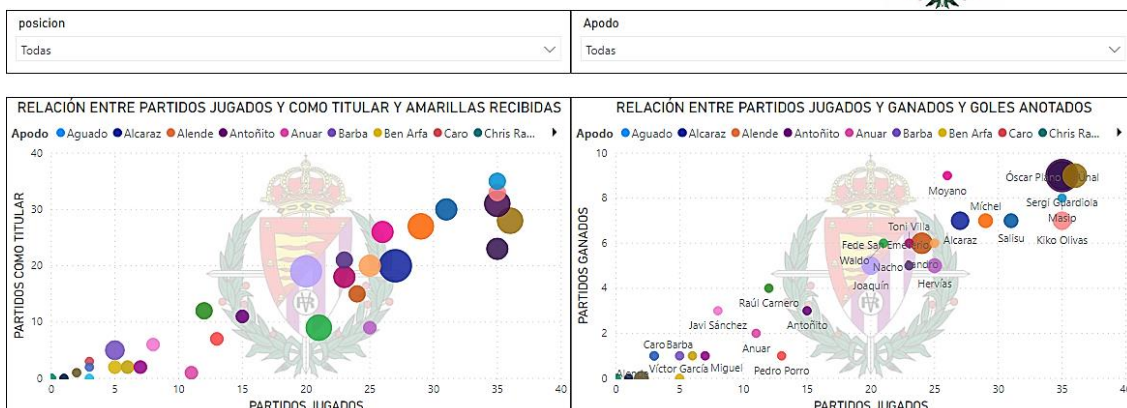


Ilustración 23. Matrices Estratégicas II. Fuente: Elaboración Propia

La sexta y última página del informe es, en varios aspectos, la página estrella. Se trata de la página en la que aparece una tabla semafórica con las estadísticas de todos los jugadores del Real Valladolid. En realidad, es la misma tabla que podemos encontrar al acudir a la página web de *BD Fútbol* (más las cuatro columnas agregadas), pero en donde cada casilla aparece en un determinado color. Verde si el dato de ese jugador para ese parámetro es bueno, amarillo si es regular y rojo si es un valor malo. También hay casillas que aparecen en azul cuando el parámetro al que se refieren no se ha tenido en cuenta a la hora de calcular la valoración global de cada jugador (en cada caso por el motivo que se explicara en el apartado oportuno).

Para desarrollar e implementar dicha tabla se ha utilizado la visualización del mismo nombre, que viene por defecto en la aplicación de escritorio de *Power BI*. Para crear la tabla, se deben ir agregando una a una, y en el orden en el que se quiere que aparezcan, todas aquellas columnas que se deseen. Una vez estén todas agregadas, se debe acudir al menú de configuración. Allí, se debe desplegar la pestaña “*Formato Condicional*”, se selecciona la columna que se quiera (se tendrá que hacer con todas aquellas que queramos que tengan el aspecto semafórico) y se activa la opción “*Color de Fondo*”. Al hacerlo, debemos pulsar sobre “*Controles Avanzados*” y se abrirá un editor. En ese editor, se elegirá la opción de dar formato por “*Reglas*”, excepto en las columnas que aparecerán en azul, que se dejará la opción por defecto, “*Escala de Colores*”.

Una vez se ha seleccionado la opción “*Reglas*”, ya se puede establecer el código de colores para esa columna. En lugar de explicar el procedimiento, que es muy sencillo, se mostrarán cuatro imágenes a modo de ejemplos de las reglas de cuatro columnas diferentes. Esas imágenes serán las correspondientes a las reglas de las columnas “*Minutos/G*”, “*Min*”, “*PS*” y “*valoracion_jugadores*”. Y las reglas son las siguientes:



Color de fondo - *Minutos/G*

Dar formato por Reglas [Más información](#)

Según el campo Suma de Minutos/G Resumen Suma

Reglas 1! Inversión del orden d... + Nueva regla

Si el valor	es mayor o igual que	400	Número	y	es menor que	5000	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	1	Número	y	es menor que	400	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	0	Número	y	es menor que	1	Número	entonces	 	↑ ↓ ×

Ilustración 24. Regla para establecer los colores de la columna "Minutos/G". Fuente: Elaboración Propia

En el caso de la columna donde se evalúan los minutos que necesita cada jugador para anotar un gol, aparecerá en color amarillo en el caso de los jugadores que tardan 400 o más minutos en anotar un gol, verde en el caso de los jugadores que tardan menos de 400 minutos y rojo en el caso de los jugadores que no han anotado ningún gol.

Color de fondo - *Min*

Dar formato por Reglas [Más información](#)

Según el campo Suma de Min Resumen Suma

Reglas 1! Inversión del orden d... + Nueva regla

Si el valor	es mayor o igual que	2000	Número	y	es menor que	5000	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	1000	Número	y	es menor que	2000	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	0	Número	y	es menor que	1000	Número	entonces	 	↑ ↓ ×

Ilustración 25. Regla para establecer los colores de la columna "Min". Fuente: Elaboración Propia

Como se puede ver, la regla para la columna de minutos jugados indica que aquellos futbolistas que hayan disputado 2000 o más minutos tendrán su casilla de color verde, los que hayan disputado entre 1000 y 1999 minutos tendrán su casilla de color amarillo y los que hayan jugado menos de 1000 minutos tendrán su casilla de color rojo.

Color de fondo - *PS*

Dar formato por Reglas [Más información](#)

Según el campo Suma de PS Resumen Suma

Reglas 1! Inversión del orden d... + Nueva regla

Si el valor	es mayor o igual que	15	Número	y	es menor o igual que	38	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	7	Número	y	es menor que	15	Número	entonces	 	↑ ↓ ×
Si el valor	es mayor o igual que	0	Porcenta	y	es menor que	7	Número	entonces	 	↑ ↓ ×

Ilustración 26. Regla para establecer los colores de la columna "PS". Fuente: Elaboración Propia

En lo que respecta a los partidos en los que el jugador parte como suplente, si estos son 15 o más su casilla aparecerá de color rojo, lo hará de color amarillo si



ha comenzado como suplente entre 7 y 14 encuentros y verde cuando ha partido desde el banquillo en menos de 7 ocasiones.

Color de fondo - *valoracion_jugadores*

Dar formato por Reglas [Más información](#)

Según el campo Suma de valoracion_jugadores Resumen Suma

Reglas ↑ Inversión del orden d... + Nueva regla

Si el valor	es mayor o igual que	18	Número	y	es menor que	100	Número	entonces	Verde	↑ ↓ ×
Si el valor	es mayor o igual que	10	Número	y	es menor que	18	Número	entonces	Amarillo	↑ ↓ ×
Si el valor	es mayor o igual que	0	Número	y	es menor que	10	Número	entonces	Rojo	↑ ↓ ×

Ilustración 27. Regla para establecer los colores de la columna "valoracion_jugadores". Fuente: Elaboración Propia

En el caso de la valoración del rendimiento global de los jugadores, dicha valoración aparecerá sobre fondo verde cuando sea igual o superior que 18, sobre fondo amarillo cuando esté entre 10 y 17 y aparecerá sobre fondo rojo cuando sea menor que 10.

Para el resto de las columnas se seguirán unos criterios semejantes a los expuestos en estos ejemplos. Una vez se ha completado toda la tabla, el resultado es el que se muestra a continuación:

	PJ	PT	PC	PS	PX	C	Min	TA	TR	G	Minutos/G	GP	GPP	GE	PG	PE	PP	minutos/GE	minutos/TA	minutos/TR	valoracion_jugadores	Valorac
	35	35	35	0	0	3	3150	2	0	0	0	0	0	39	8	15	12	80,77	1.575,00	0,00	27	★★★★
	35	33	32	2	1	2	2941	2	0	2	1470	0	0	0	7	14	14	0,00	1.470,50	0,00	27	★★★★
	31	30	28	1	2	3	2724	4	0	1	2724	0	0	0	7	14	10	0,00	681,00	0,00	25	★★★★
iones)	35	31	18	4	13	0	2659	6	0	8	332	0	0	0	9	13	13	0,00	443,17	0,00	24	★★★★
	35	23	17	12	6	1	2340	4	0	6	390	2	0	0	9	13	13	0,00	585,00	0,00	24	★★★★
iones)	36	28	15	8	13	1	2507	6	0	4	626	0	0	0	9	15	12	0,00	417,83	0,00	19	★★★★
	26	26	22	0	4	8	2229	4	0	0	0	0	0	0	9	11	6	0,00	557,25	0,00	19	★★★★
	23	21	18	2	3	6	1894	2	0	0	0	0	0	0	5	11	7	0,00	947,00	0,00	19	★★★★
	29	27	13	2	14	4	2186	6	0	1	2186	1	0	0	7	12	10	0,00	364,33	0,00	18	★★★★
	27	20	16	7	4	3	1927	10	0	2	963	0	0	0	7	10	10	0,00	192,70	0,00	17	★★★★
	25	20	14	5	6	7	1753	4	0	0	0	0	0	0	6	10	9	0,00	438,25	0,00	16	★★★★
	12	12	10	0	2	4	1009	2	0	0	0	0	0	0	4	3	5	0,00	504,50	0,00	15	★★★★
iones)	15	11	9	4	2	12	1030	1	0	0	0	0	0	0	3	6	6	0,00	1.030,00	0,00	15	★★★★
	24	15	8	9	10	2	1302	2	0	3	434	0	0	0	6	9	9	0,00	651,00	0,00	15	★★★★
	20	19	16	1	3	8	1660	9	0	2	830	0	0	0	5	8	7	0,00	184,44	0,00	15	★★★★
	25	9	1	16	8	8	1015	1	0	1	1015	0	0	0	5	11	9	0,00	1.015,00	0,00	13	★★★
iones)	2	1	0	1	1	4	88	0	0	1	88	0	0	0	0	0	2	0,00	0,00	0,00	12	★★★
	23	18	2	5	16	12	1357	4	0	0	0	0	0	0	6	10	7	0,00	339,25	0,00	11	★★★
	13	7	8	6	1	10	737	1	0	0	0	0	0	0	1	5	7	0,00	737,00	0,00	10	★★★
	8	6	4	2	2	7	501	1	0	0	0	0	0	0	3	1	4	0,00	501,00	0,00	10	★★★
iones)	3	0	0	3	0	2	62	0	0	0	0	0	0	0	1	1	1	0,00	0,00	0,00	10	★★★
	6	2	1	4	1	1	217	1	0	0	0	0	0	0	1	2	3	0,00	217,00	0,00	10	★★★
	5	5	5	0	0	4	450	3	0	0	0	0	0	0	1	0	4	0,00	150,00	0,00	9	★★★
	5	2	0	3	2	5	159	1	0	0	0	0	0	0	0	3	2	0,00	159,00	0,00	9	★★★
	11	1	0	10	1	1	183	1	0	0	0	0	0	0	2	5	4	0,00	183,00	0,00	9	★★★
	3	2	1	1	1	4	166	0	0	0	0	0	0	0	1	0	2	0,00	0,00	0,00	9	★★★
	3	3	2	0	1	8	245	0	0	0	0	0	0	3	1	0	2	81,67	0,00	0,00	9	★★★
	544	418	292	126	126	178	37623	84	0	31	11058	3	0	43	130	212	202	187,44	13.718,56	0,00	449	★★★★

Tabla 7. Tabla semafórica de comparación del rendimiento de los jugadores del Real Valladolid. Fuente: Elaboración Propia

No aparecen las columnas con el dorsal, el apodo y la posición de cada jugador debido a que no cabía todo en la pantalla al mismo tiempo. Tampoco cabían



todas las filas. Para movernos por la tabla, se pueden usar las barras de desplazamiento, tanto la vertical como la horizontal.

Justo encima de la tabla aparece un selector donde se puede filtrar por posiciones. De nuevo, insistiendo en la idea que ya se ha expuesto anteriormente, la utilidad de este selector no está en esta tabla, si no en la que aparecerá cuando se cargue la macro-tabla con las estadísticas de todos los jugadores de todos los equipos de primera, segunda y segunda división B de España y de las primeras divisiones de Alemania, Inglaterra, Italia, Francia, Portugal y Holanda.

En la parte superior de la página, justo debajo del encabezado, aparecen cuatro etiquetas que se han agregado a través de la visualización “*Tarjeta de Varias Filas*”. Pasándolas los campos correspondientes, estas etiquetas muestran al jugador que mayor valoración ha obtenido en su posición, pues una tarjeta está filtrada para porteros, otra para defensas, otra para centrocampistas y otra para delanteros. Además, cada una dispone de una barra de desplazamiento que permite bajar para ir viendo más jugadores, en orden decreciente de valoración obtenida. A modo de ejemplo, en la siguiente ilustración se muestra, de forma desplegada, la tarjeta de clasificación de los centrocampistas:

18	★★★★☆		Michel
17	★★★★☆		Alcaraz
16	★★★★☆		Fede San Emeterio
13	★★★★☆		Hervías
11	★★★★☆		Toni Villa
10	★★★★☆		Kike Pérez

Ilustración 28. Tarjeta donde aparecen los centrocampistas del Real Valladolid, en orden decreciente de valoración obtenida. Fuente: Elaboración Propia

Con todo lo expuesto, ya estaría completa la sexta página del informe, con la cual se daría por concluido éste. El aspecto general de esta página es el que se muestra en la siguiente ilustración:



TABLA SEMAFÓRICA Y COMPARATIVA DE VALORACIÓN POR PUESTOS



Portero más valorado
27 ★★★★★
Masip

Defensa más valorado
27 ★★★★★
Kiko Olivas

Centrocampista más valorado
18 ★★★★★
Michel

Delantero más valorado
24 ★★★★★
Sergi Guardiola

Estadísticas por posición

#	Apellido	posición	PJ	PT	PC	PS	PX	C	Min	TA	TR	G	Minutos/G	GP	GPP	GE	PG	PE	PP	minutos/GE	minutos/TA	minutos/TR	valoración Jugadores	Valoración	
1	Masip	Portero	35	35	33	0	0	3	3150	2	0	0	0	0	0	0	39	6	15	15	80,77	1.575,00	0,00	27	★★★★★
4	Kiko Olivas	Central	35	33	32	2	1	2	2941	2	0	2	1470	0	0	0	7	14	15	0,00	1.470,00	0,00	27	★★★★★	
27	Salsu	Central	31	30	26	1	2	3	2724	4	0	1	2724	0	0	0	7	14	10	0,00	681,00	0,00	25	★★★★★	
7	Sergi Guardiola	Delantero (Varias Posiciones)	35	31	16	4	3	0	2659	6	0	8	332	0	0	0	9	13	13	0,00	443,17	0,00	24	★★★★★	
9	Únal	Delantero Centro	35	23	17	12	6	1	2340	4	0	6	390	2	0	0	9	13	13	0,00	585,00	0,00	24	★★★★★	
10	Oscar Plano	Delantero (Varias Posiciones)	36	28	15	8	1	1	2507	6	0	4	626	0	0	0	9	15	15	0,00	417,83	0,00	19	★★★★★	
17	Moyano	Lateral Derecho	26	26	22	0	4	6	2229	4	0	0	0	0	0	0	8	11	6	0,00	557,25	0,00	19	★★★★★	
22	Nacho	Lateral Izquierdo	23	21	18	2	3	6	1894	2	0	0	0	0	0	0	5	11	7	0,00	947,00	0,00	19	★★★★★	
21	Michel	Centrocampista	29	27	13	2	4	4	2186	6	0	1	2186	1	0	0	7	12	10	0,00	364,33	0,00	17	★★★★★	
14	Alcaraz	Centrocampista	27	20	16	7	4	3	1927	10	0	2	963	0	0	0	7	10	10	0,00	182,70	0,00	17	★★★★★	
16	Fede San Emeterio	Centrocampista	25	20	14	5	6	7	1753	4	0	0	0	0	0	0	6	10	9	0,00	438,25	0,00	16	★★★★★	
6	Raúl Carrero	Lateral Izquierdo	12	12	10	0	2	4	1009	2	0	0	0	0	0	0	4	5	3	0,00	504,50	0,00	15	★★★★★	
18	Antoñito	Defensa (Varias posiciones)	15	11	9	4	2	12	1030	1	0	0	0	0	0	0	3	6	6	0,00	1.050,00	0,00	15	★★★★★	
20	Sandro	Delantero (Varias Posiciones)	24	15	10	9	1	3	1302	2	0	3	434	0	0	0	6	9	9	0,00	851,00	0,00	15	★★★★★	
24	Joaquín	Central	20	19	16	1	3	2	1660	3	0	2	830	0	0	0	5	8	7	0,00	164,44	0,00	15	★★★★★	
11	Henriás	Centrocampista	25	9	16	16	8	3	1015	1	0	1	1015	0	0	0	5	11	9	0,00	1.315,00	0,00	13	★★★★★	
35	Víctor García	Delantero (Varias Posiciones)	1	1	1	1	1	4	88	0	0	1	88	0	0	0	0	0	0	0,00	0,00	0,00	12	★★★★★	
19	Toni Villa	Centrocampista	23	16	13	3	16	4	1357	4	0	0	0	0	0	0	6	10	7	0,00	339,25	0,00	11	★★★★★	
Total			544	418	292	126	126	178	37623	84	0	31	11058	3	0	43	130	212	202	187,44	13.718,56	0,00	449	★★★★★	

Ilustración 29. Vista de la página donde aparece la tabla semafórica y las tarjetas de jugadores mejor valorados por posición. Fuente: Elaboración Propia

Una vez desarrollada completamente la explicación sobre la creación e implementación del informe sobre el análisis del rendimiento de los jugadores del Real Valladolid, y según se expuso al principio de este subapartado, se comentarán las diferencias que se han introducido en el segundo informe, el de análisis y comparación de los jugadores del resto de los equipos.

En casi todas las páginas, el resultado visual es muy parecido. La única que escapa de esta afirmación es la portada, que ha cambiado la imagen de fondo y los colores para tener un aspecto que la diferencia de la otra:



Ilustración 30. Portada del informe sobre el análisis de los jugadores del resto de los equipos. Fuente: Elaboración Propia



En lo que respecta al comparador de los jugadores de campo sólo se han introducido dos modificaciones. En primer lugar, de la misma forma que se introdujo la imagen con el rostro del jugador seleccionado, se ha introducido la imagen del escudo del equipo indicado, utilizando la visualización “Simple Image”.

Para obtener el escudo del equipo se ha procedido de la siguiente forma. En primer lugar, usando como referencia al Real Valladolid, se buscó el enlace web en el que aparece la imagen del escudo del equipo. Es el siguiente:

<https://www.bdfutbol.com/i/eg/37.png>

Como se puede apreciar, el único elemento que cambiará de un enlace a otro es el código del equipo, mientras que el resto de la estructura de la dirección URL permanecerá fija. Esto debería recordarnos al momento en el que se trató de generar la macro-tabla, parametrizando una parte del enlace web que llevaba a las tablas de estadísticas de los jugadores de cada uno de los equipos. Aquí, por consiguiente, podemos proceder del mismo modo. Para ello, se ha agregado, al lado de la columna donde aparecen los códigos de los clubes la siguiente columna personalizada:

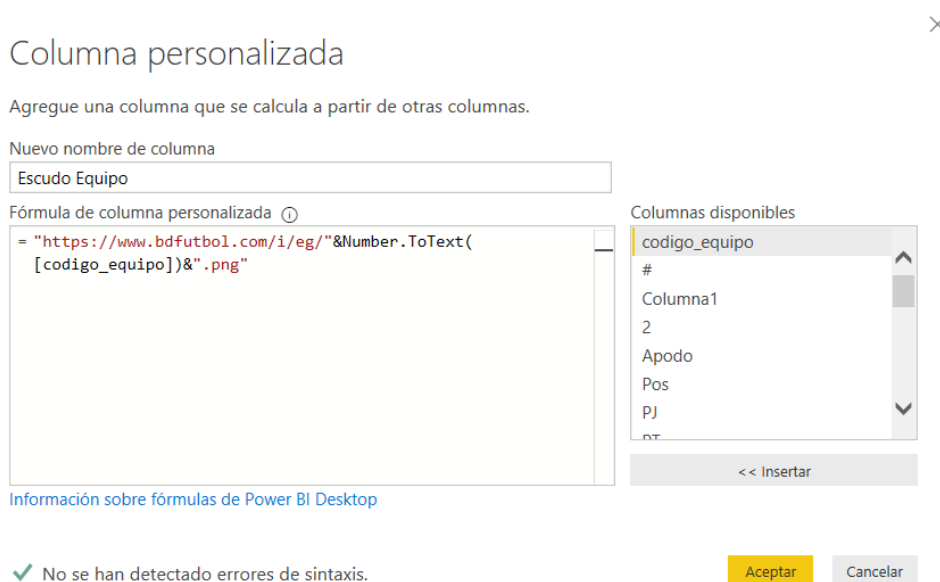


Ilustración 31. Primera columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia

Sin embargo, aquí surge una diferencia con respecto al caso de la macro-tabla, y es que no queremos obtener datos y tablas y tampoco basta con obtener un enlace donde aparece la imagen, sino que se necesita generar la imagen en el programa para poder mostrarla en el informe. Para ello, ha sido necesario crear las siguientes dos columnas:



Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna
escudo_equipo_2

Fórmula de columna personalizada ⓘ
= Web.Contents([escudo_equipo])

Columnas disponibles
codigo_equipo
nombre_equipo
escudo_equipo

<< Insertar

Información sobre fórmulas de Power BI Desktop

✓ No se han detectado errores de sintaxis. Aceptar Cancelar

Ilustración 32. Segunda columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia

La función “Web.Contents” devuelve, como valor binario, el contenido que se ha descargado de una determinada dirección URL.

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna
Escudo del Equipo

Fórmula de columna personalizada ⓘ
= "data:image/jpeg;base64, "&Binary.ToText([escudo_equipo_2], BinaryEncoding.Base64)

Columnas disponibles
codigo_equipo
nombre_equipo
escudo_equipo
escudo_equipo_2

<< Insertar

Información sobre fórmulas de Power BI Desktop

✓ No se han detectado errores de sintaxis. Aceptar Cancelar

Ilustración 33. Tercera columna agregada para obtener la imagen con el escudo de los distintos equipos. Fuente: Elaboración Propia

Aquí ya se genera dicha imagen, tratando el contenido como valores binarios (por lo hecho a través de la columna agregada anteriormente), estableciendo el formato en el que aparecerá la imagen y posibilitando la visualización de éstas en el informe.

La segunda modificación que se ha introducido en el comparador de jugadores de campo, con respecto al del informe del Real Valladolid, es la agregación de dos selectores más para cada una de las dos fichas. Es decir, cuatro selectores más de segmentación de datos. Puesto que buscar el nombre del jugador que queremos evaluar, entre miles de futbolistas, sería una tarea que nos llevaría varios



minutos, se han introducido dos selectores más, para poder filtrar previamente tanto por equipo, como por posición. Para hacer esto posible, ha sido necesario crear una columna que asociara el nombre del equipo con su código. El proceso ha sido el siguiente:

Lo primero que se ha hecho es ver dónde se podía encontrar el nombre del equipo en la página donde aparecía la tabla con las estadísticas de los jugadores. Para ello, se ha procedido como si se fuera a agregar una nueva tabla:

Obtener datos → *Web* → *Dirección URL* → *Aceptar* → *Vista Web* → *Agregar tabla mediante ejemplos*

Una vez nos encontramos en este punto, comenzamos a escribir el nombre del equipo en la primera fila, hasta que *Power BI* lo reconozca y localice dónde se encuentra. Lo seleccionamos y aceptamos, obteniendo el siguiente resultado:

De web

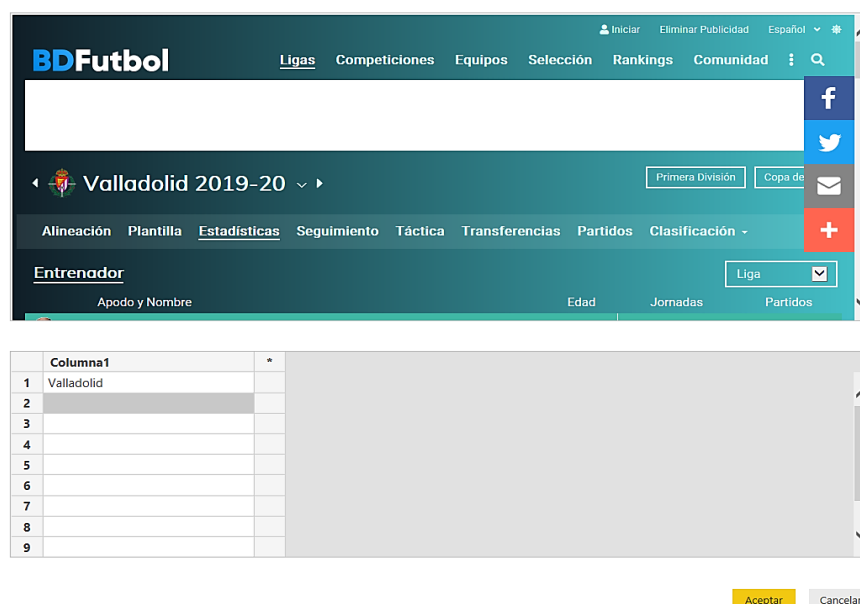


Ilustración 34. Primer paso llevado a cabo para asociar el código de un equipo con su nombre. Fuente: Elaboración Propia

Dando a “*Aceptar*”, se carga una tabla con una única fila donde aparece el nombre “*Valladolid*”, que es la forma en la que *BD Fútbol* ha registrado al Real Valladolid CF.

El siguiente paso consiste en editar la tabla agregada para que, en lugar de leer el nombre del Real Valladolid del enlace en el que se encuentra, lea el nombre de todos los equipos iterando en función de la columna donde viene el código de cada uno. Tal y como se hizo a la hora de generar la macro-tabla con las estadísticas de todos los jugadores. Para ello, se abre el “*Editor Avanzado*” situado en el “*Editor de Power Query*” y, mediante el uso de la función “*Number.ToText*”, se logra la



iteración deseada. En la siguiente imagen se puede ver la pantalla del editor una vez se han programado las modificaciones pertinentes:



Ilustración 35. Pantalla del Editor Avanzado de Power Query donde se ha programado la iteración para obtener el nombre de cada equipo. Fuente: Elaboración Propia

Por último, en la tabla donde aparecen las estadísticas de todos los jugadores, se crea la columna personalizada en la que debe aparecer el nombre de cada equipo al lado del código que *BD Fútbol* asigne a cada uno.

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

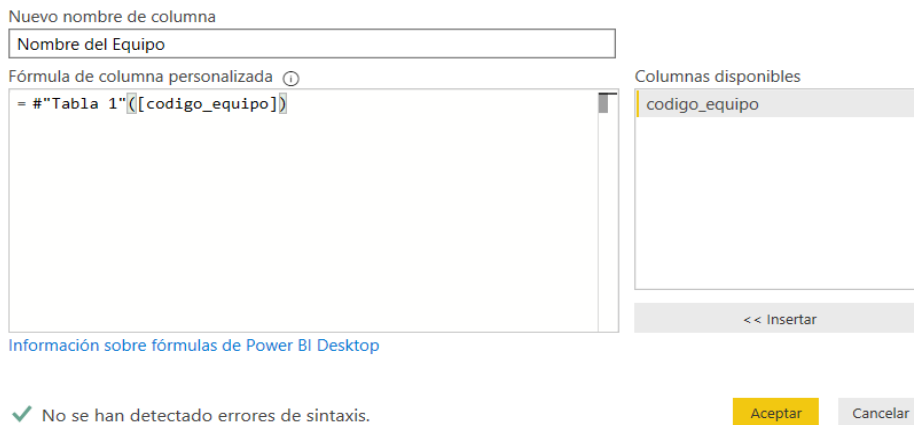


Ilustración 36. Columna agregada para mostrar el nombre del equipo. Fuente: Elaboración Propia

Una vez creada la columna, se expande y se cambia el tipo de datos que contiene a *Texto*. De esta forma, ya estaría asociado cada nombre al código del equipo y ya se puede agregar ese campo al filtro correspondiente del comparador de jugadores.

Hecho todo esto, ya estaría finalizado el comparador de jugadores de campo. En la siguiente ilustración se puede ver el resultado obtenido, en el que, a modo de



ejemplo, se compara al jugador del FC Barcelona Leo Messi con el jugador del Bayern de Múnich Robert Lewandowski:



Ilustración 37. Comparador de jugadores del informe de análisis de todos los equipos. Fuente: Elaboración Propia

En lo que respecta al comparador de porteros, se ha procedido, como se hizo en el informe anterior, a duplicar esta página y modificar las cosas que correspondan. En este caso, a mayores de las modificaciones que ya se comentaron en el anterior supuesto, se ha eliminado el selector de posición, puesto que sólo estará habilitada la demarcación de portero. Por lo demás, el resultado visual es el mismo. En la siguiente ilustración se compara al portero del Real Madrid Thibaut Courtois con el guardameta del FC Barcelona Marc-André ter Stegen:

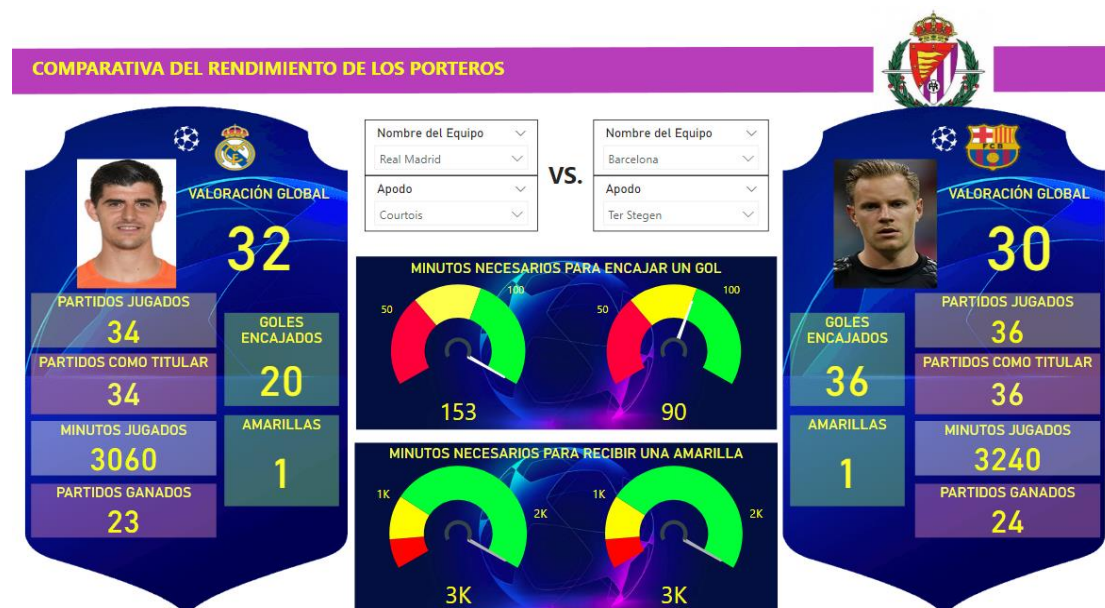


Ilustración 38. Comparador de porteros del informe de análisis de todos los equipos. Fuente: Elaboración Propia



En cuanto a las páginas cuatro y cinco del informe, las relativas a las matrices estratégicas y los gráficos de dispersión, se han mantenido exactamente igual que en el informe de análisis de la plantilla del Real Valladolid, con la excepción de que se ha añadido un selector más, el que permite filtrar por el nombre del equipo, que se suma a los que ya había de posición y apodo o nombre del jugador. Como ya se comentó anteriormente, ahora esos selectores de segmentación de datos cobran una notable importancia, pues representar en una misma visualización a miles de jugadores es inviable, ya que sólo se vería un gigante cúmulo de círculos en el que sería imposible distinguir nada.

En la siguiente imagen se puede ver la cuarta página del informe, en la que las matrices estratégicas han sido filtradas para mostrar únicamente los círculos correspondientes a los centrocampistas y delanteros del Club Atlético de Madrid, del FC Barcelona y del Real Madrid CF:



Ilustración 39. Vista filtrada de la página "Matrices Estratégicas I" del segundo informe. Fuente: Elaboración Propia

Por su parte, en la quinta página se han filtrado los gráficos de dispersión para que únicamente se muestren los círculos correspondientes a todos los jugadores del Club Atlético de Madrid, del FC Barcelona y del Real Madrid, excepto los porteros y los futbolistas que no tengan ninguna posición asignada. El resultado es el que se puede ver en la siguiente ilustración:

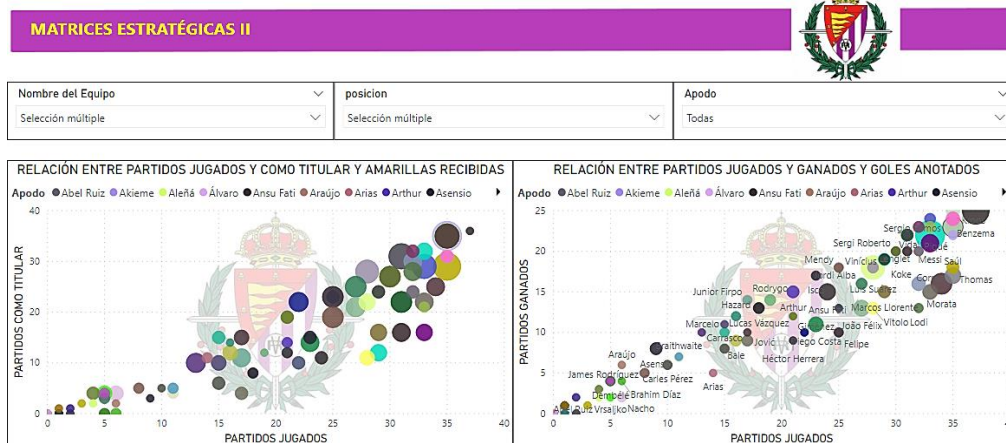


Ilustración 40. Vista filtrada de la página "Matrices Estratégicas II" del segundo informe. Fuente: Elaboración Propia

Por último, la página que cerraba el informe sobre el análisis de los jugadores del Real Valladolid, la de la tabla semafórica de comparación, se ha desdoblado en dos páginas. La primera de ellas, que es la que se considera homóloga a la del otro informe, ha cambiado ligeramente su aspecto para ofrecer más información, puesto que ahora se puede evaluar a más equipos. Se ha añadido un selector más, para poder filtrar por el nombre del equipo y se han incluido unas etiquetas donde se indica la valoración global del equipo en su conjunto y en función de cada una de las demarcaciones. También se ha incluido el escudo del equipo por el que se filtra.

Por otro lado, la séptima, y última, página es un comparador de equipos en el que se pueden ver al mismo tiempo y en el mismo panel las valoraciones globales de los equipos en su conjunto y en función de cada posición. Es decir, igual que lo de la página anterior, pero poniendo frente a frente a dos clubes, lo cual facilita las labores de análisis y comparación.

Para desarrollar todo esto, ha sido necesario obtener el código que *BD Fútbol* asigna a cada jugador para poder diferenciar a aquellos que tienen el mismo nombre y evaluarlos por separado.

Para obtener el código de cada jugador se ha tenido que leer la dirección URL a la que lleva *BD Fútbol* cuando pulsamos sobre el apodo del jugador en la tabla de estadísticas de los futbolistas de cada equipo, grabar dicha dirección en una columna y extraer el código. Por ejemplo, en el caso del guardameta del Real Valladolid Jordi Masip, la dirección URL que lleva a su ficha es la siguiente:

<https://www.bdfutbol.com/es/j/i300050.html>

Como se puede intuir, el código del jugador es "300050" y el resto de la dirección será la misma para cualquier futbolista que esté registrado en la página web. Por consiguiente, se debe recoger este enlace web y aislar el código para poder insertarlo en una nueva columna. Para ello, se han tenido que crear las dos columnas cuya programación se muestra en las siguientes imágenes:

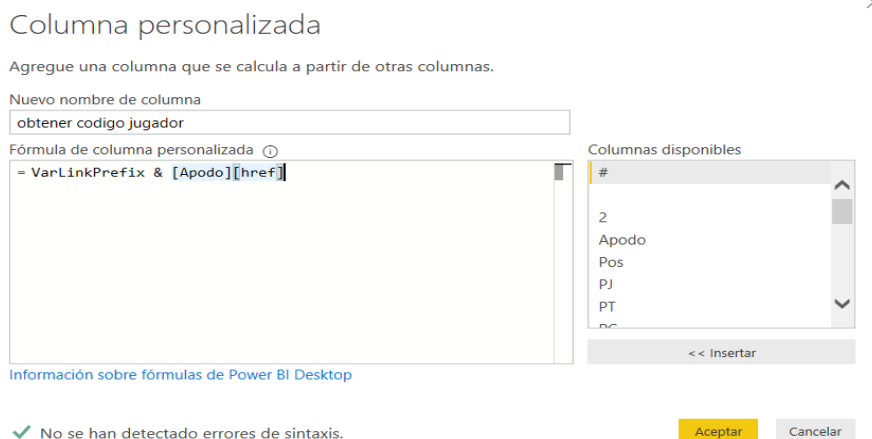


Ilustración 41. Primera columna agregada para obtener el código de cada jugador. Fuente: Elaboración Propia

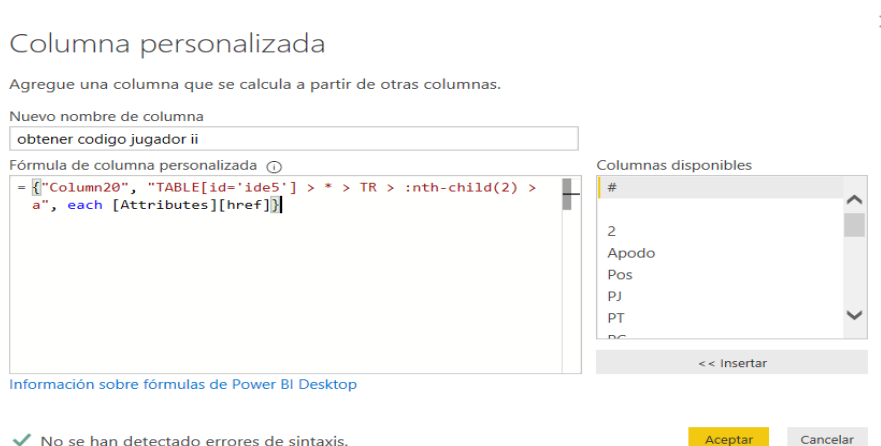


Ilustración 42. Segunda columna agregada para obtener el código de cada jugador. Fuente: Elaboración Propia

Sin embargo, es necesario resaltar que este paso se hizo al mismo tiempo que se cargaban en el programa las imágenes de los jugadores, debido a que estas se encuentran en la ficha que cada jugador tiene en su enlace correspondiente de *BD Fútbol*. Posteriormente, las columnas auxiliares que se crearon como pasos intermedios fueron eliminadas, dejando únicamente aquellas necesarias para la elaboración y confección del informe.

Para evitar que pudiera haber problemas al crear una relación entre la columna “Apodo” de la tabla de estadísticas de los jugadores y la columna “Apodo_jugador” de la tabla donde se han cargado las imágenes de los jugadores y su código o identificador, también se ha creado una columna, en esta última tabla, donde se ha agrupado el nombre del jugador con su identificador.

Con todo lo expuesto hasta ahora, sólo queda mostrar el resultado obtenido, el cual se puede ver a continuación. En primer lugar, se mostrará una imagen donde se aprecia la sexta página, la de la tabla semafórica y comparativa de valoración por puestos. En este caso, se evalúa al Sevilla FC:

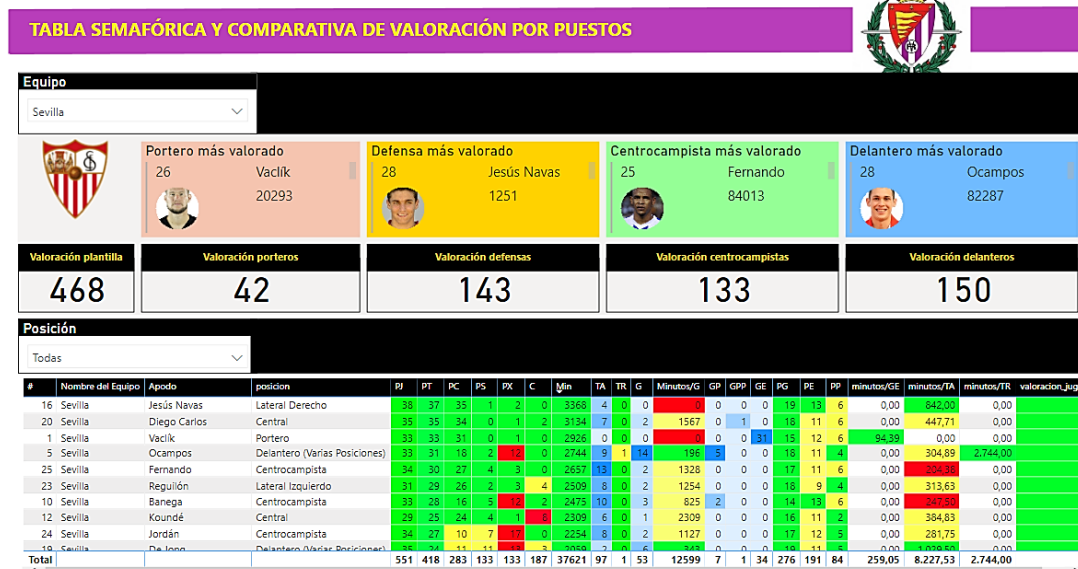


Ilustración 43. Tabla semafórica y comparativa de valoración por puestos del segundo informe. Fuente: Elaboración Propia

Asimismo, la página de comparación de equipos queda como se puede ver seguidamente, donde se compara al Real Valladolid con el Deportivo Alavés:



Ilustración 44. Tabla comparativa de equipos según su valoración por puestos. Fuente: Elaboración Propia

Una vez hecho esto, se da por terminado el informe sobre el análisis del rendimiento de los jugadores de todos los equipos.





6. RESULTADOS OBTENIDOS Y CONCLUSIONES

6.1. VALORACIÓN DE LOS RESULTADOS OBTENIDOS Y DEL GRADO DE CONSECUCCIÓN DE LOS OBJETIVOS MARCADOS

Al principio del presente Trabajo de Fin de Grado, en el apartado de “*Introducción y Objetivos*”, se mencionaron una serie de objetivos o metas que se pretendían alcanzar mediante el desarrollo de este trabajo. En este punto, ha llegado el momento de comentar y evaluar brevemente hasta que punto se han conquistado dichas metas. Se comenzará por los denominados objetivos complementarios.

En primer lugar, se habló de conocer qué son el Big Data y el análisis de datos y cuáles son las semejanzas, características y diferencias que los identifican. Si bien es cierto que se han estudiado las posibles definiciones para cada uno de los conceptos, así como las principales propiedades y particularidades de cada uno, también es justo reconocer que no es posible adentrarse en profundidad en unos campos tan inmensos, complejos y en plena evolución como los que se han tratado en este trabajo. Por ello, se puede considerar como superado este objetivo, pero teniendo presente que sólo se ha estudiado la parte visible de este gran iceberg.

En segundo lugar, se comentó la importancia de comprender y analizar los debates que existen en la actualidad con respecto al Big Data y al análisis de datos, así como proponer algunas líneas de evolución. Sí que se han relatado los problemas que existen con relación a la privacidad de los usuarios cuando éstos navegan por el mundo digital y otras cuestiones de carácter ético y social que están presentes hoy en día, como la desigualdad que hay entre los países y las sociedades más ricas con respecto a la población de los países subdesarrollados y en vías de desarrollo a la hora de avanzar en la implementación de las herramientas de Big Data. En lo que respecta a las líneas de evolución futura, sí que se ha hablado de varios campos donde estas técnicas y herramientas irán ganando más peso con el paso de los años, como la medicina, la agricultura o el periodismo, pero siguiendo el hilo conductor del TFG, donde se ha hecho especial hincapié es en el campo de la gestión deportiva.

Lo tercero que se propuso fue analizar cómo podían mejorar la vida de las personas y la evolución de la sociedad las técnicas y herramientas de Big Data y análisis de datos. Este objetivo se ha desarrollado y estudiado de forma conjunta con el anterior, pues guarda una relación directa la propuesta de campos donde estas herramientas puedan facilitar y mejorar las cosas con la influencia que tengan en el mundo real.

El cuarto punto de los objetivos complementarios trataba sobre el estudio de la Inteligencia de Negocio o Inteligencia Empresarial. Se ha considerado que no era necesario adentrarse tanto teóricamente en este punto como en los anteriores, debido a que se trataría de una forma operativa en el supuesto práctico y que en el grado de Ingeniería en Organización Industrial se trata de un modo mucho más exhaustivo que los conceptos anteriores. Pese a ello, sí que se han estudiado



convenientemente las distintas posibles definiciones de la Inteligencia de Negocio, así como sus características, ventajas e inconvenientes y diferencias y relaciones con el análisis de datos y el Big Data.

Por último, en quinto lugar, se mencionaba la imperiosa necesidad de conocer el programa con el que se iba a desarrollar la herramienta de análisis en la gestión deportiva: *Power BI*. Antes de adentrarse en el supuesto práctico, se ha desarrollado una breve exposición teórica sobre los distintos componentes del programa, así como sus características y ventajas. A modo complementario, se han mencionado y descrito brevemente otros programas y herramientas de características similares. Sin embargo, ha sido el caso práctico y el desarrollo de los informes los que han provocado que se adquieran unos más que aceptables conocimientos sobre el funcionamiento de este programa. Es necesario explicar que este TFG no es un manual de uso de *Power BI*, pero a mí sí que me ha tocado aprender a utilizarlo, por lo que hay un gran número de horas dedicadas, tanto a aprender a utilizar el programa como a resolver diversos problemas que han ido surgiendo a lo largo del trabajo, que no tienen reflejo en esta memoria.

Y, a mayores de los objetivos complementarios, se encuentra el principal propósito que se ha pretendido alcanzar con este TFG. Éste es la implementación, mediante *Power BI*, de una herramienta de análisis en la gestión deportiva. El resultado son los dos informes generados:

- Herramienta de Análisis en la Gestión Deportiva – Parte de Análisis del rendimiento de los jugadores del Real Valladolid.
- Herramienta de Análisis en la Gestión Deportiva – Parte de Análisis del rendimiento de los jugadores de todos los equipos.

Estos informes han sufrido muchas variaciones, cambios y modificaciones desde el comienzo de este trabajo, hasta dar con un resultado que, a título personal, se puede considerar como más que satisfactorio. Se ha dado prioridad a aspectos que se consideran fundamentales para alcanzar la satisfacción del cliente. Algunos de ellos son los siguientes:

- Facilidad para desplazarse a través de la herramienta.
- Facilidad para trabajar y operar con los informes.
- Aspecto atractivo y agradable de los paneles.
- Visualizaciones oportunas para ayudar en la toma de decisiones.
- Elementos que marquen la diferencia con respecto a otros comparadores y páginas web de propósitos similares ya existentes.

6.2. CONCLUSIONES EXTRAÍDAS DE LA ELABORACIÓN DE ESTE TRABAJO

La primera conclusión que se puede valorar ya ha sido parcialmente comentada en el apartado anterior. El Big Data y el análisis de datos son dos campos de unas magnitudes inabarcables para un trabajo de estas características. Se ha llevado a cabo una labor de investigación y documentación bastante exhaustiva, pero



aún así, no se ha podido profundizar en todos los conceptos que se han tratado. Además, otra conclusión a la que se ha llegado fácilmente es que el Big Data y el análisis de datos no son materias que únicamente estén relacionadas con el mundo de la ingeniería en organización y de la ingeniería industrial, sino que también guardan una estrecha relación con la informática y la estadística, entre otras áreas.

Otra idea que, en línea con lo anterior, se ha extraído de la parte teórica de esta memoria es que las técnicas y herramientas de tratamiento de datos no son solo aplicables en campos de la ingeniería, la ciencia y la tecnología, sino que prácticamente cualquier sector de la sociedad se puede ver beneficiado de la incorporación de éstas. Se ha visto que la medicina puede avanzar en materia de localización de brotes de enfermedades y de diagnóstico precoz de patologías, entre otras posibilidades, que la agricultura puede mejorar en aspectos como la optimización del uso del suelo o que en el deporte se pueden prevenir lesiones, mejorar la planificación deportiva u optimizar la gestión de los recursos económicos de los clubes.

La tercera conclusión que merece la pena comentar está relacionada con los debates que actualmente están sobre la mesa en materia de tratamiento de la información. Queda patente que el mayor reto que afrontarán los profesionales del Big Data y del análisis de datos en los próximos años consiste en desarrollar e implementar todas las ventajas que ofrecen a los ciudadanos estos conceptos, sin que la privacidad y la seguridad en línea se vean afectadas. Obtener publicidad personalizada, evolucionar el Internet de las Cosas, facilitar la generación y el análisis de estudios a las empresas, etc. son cosas que a todos pueden beneficiar, pero se volverán negativas si esos beneficios se producen a costa de la vulneración de algunos de los derechos más importantes de los disponen las personas.

Otra idea importante que refleja este trabajo es que el Big Data no sustituye o reemplaza al análisis de datos, sino que lo complementa. Y lo mismo sucede con la relación entre el Big Data y la Inteligencia de Negocio o Inteligencia Empresarial. Tal y como se ha comentado en los apartados correspondientes, no todas las organizaciones necesitan disponer de avanzadas herramientas de Big Data, sino que, en la mayoría de las empresas, el volumen de datos con el que necesitan trabajar es relativamente pequeño. Por ello, en muchas ocasiones, será suficiente con aplicar algunos principios de la inteligencia de negocio para ver cómo los resultados obtenidos son más que satisfactorios.

Centrándonos ahora en el caso práctico planteado, y siguiendo lo ya comentado en el apartado anterior, se puede afirmar que es fundamental desarrollar un proyecto teniendo como perspectiva el resultado que se le quiere ofrecer al cliente. En ese sentido, cada vez cobra más importancia todo lo relacionado con la visualización y la representación de los resultados; y prueba de ello es la existencia de programas como *Power BI* o como algunos de los comentados en el apartado correspondiente, donde lo visual destaca por encima del resto.



Tal y como se explicó al principio del caso práctico, y en consonancia con lo que se acaba de exponer, se puede concluir que es absolutamente imprescindible conocer muy bien al cliente y estudiar detalladamente los requisitos que nos solicita si se quiere llevar el proyecto a buen puerto. En caso contrario, el fracaso estará casi garantizado.

Otra conclusión que se puede extraer de la parte práctica de este trabajo, posiblemente la más importante, es que, no disponer de la base de datos de la que se nutre a la herramienta, constituye una limitación muy relevante en muchos aspectos. En primer lugar, sólo se puede evaluar a los futbolistas en función de los parámetros que mide dicha base de datos. En segundo lugar, se pierde mucho tiempo en comprender el funcionamiento y la estructura de la BBDD y, posteriormente, en resolver los problemas que surgen a la hora de cargar las tablas, leer las direcciones URL... y junto a estos dos ejemplos, un sinfín de limitaciones más. Con total seguridad, se puede afirmar que haber dispuesto de la base de datos habría facilitado enormemente el desarrollo e implementación de la herramienta de análisis en la gestión deportiva.

Posiblemente se puedan extraer muchas más ideas y conclusiones de este trabajo. Sin embargo, se considera que las más importantes y generales son las que se han expuesto aquí, por lo que, con la anterior valoración, se da por finalizada esta memoria.



7. BIBLIOGRAFÍA

- Alexis Rendón, Y. (2019). *Bases de datos relacionales vs. no relacionales*. Academia Pragma. <https://www.pragma.com.co/academia/lecciones/bases-de-datos-relacionales-vs.-no-relacionales>
- Antón Carranza, M. (2017). *Identificación del talento en la Organización: El Big Data aplicado al fútbol*. <http://uvadoc.uva.es/handle/10324/27451>
- ARISER. (2015). *Historia del Big Data – Del comienzo del análisis de datos a nuestros días*. Big Data para curiosos. <https://bigdataparacuriosos.wordpress.com/historia-big-data/>
- Bandrés Balet, C. (2019). *El origen y la evolución de la movilidad de las personas con la utilización de las nuevas tecnologías*. <http://uvadoc.uva.es/handle/10324/37435>
- BBVA. (2017). *Las cinco uves del big data*. BBVA. <https://www.bbva.com/es/las-cinco-uves-del-big-data/>
- BBVA. (2018). *¿Qué hace el «big data» en el mundo del deporte?* BBVA. <https://www.bbva.com/es/que-hace-el-big-data-en-el-mundo-del-deporte/>
- Benítez, R., Escudero, G., & Kanaan, S. (2013). *Inteligencia artificial avanzada*.
- Berlanga Silvente, V., José Rubio Hurtado, M., Vilà Baños, R., Berlanga Silvente Profesora, V., José Rubio Hurtado Profesora, M., & Vilà Baños Profesora, R. (2013). *Cómo aplicar árboles de decisión en SPSS*. *diposit.ub.edu*, 6(1), 65-79. <https://doi.org/10.1344/reire2013.6.1615>
- Big Data International Campus. (2016). *Big Data en el deporte*. Big Data International Campus. <https://www.campusbigdata.com/big-data-blog/item/116-big-data-en-el-deporte>
- Big Data International Campus. (2020). *¿Para qué sirve el Big data?* Big Data International Campus. <https://www.campusbigdata.com/big-data-blog/item/77-bigdata-macrodatos-5v-datascience>
- Blanco, L. (2016). *El efecto Moneyball*. INUSUAL. <https://inusual.com/es/blog/el-efecto-moneyball>
- Bock, T. (2018). *What is Raw Data?* Displayr. <https://www.displayr.com/what-is-raw-data/>
- BusinessGoOn. (2019). *HISTORIA DEL BIG DATA*. <https://businessgoon.com/historia-big-data/>
- Carlos, J., & Castilla, R. (2006). *La actuación policial frente a los déficits de seguridad de Internet*. *dialnet.unirioja.es*. www.uoc.edu/idp
- Castro Franco, M., & Domenech, M. B. (2014). *Agro Big Data: el próximo desafío*.
- Checa Cruz, M. (2019). *El Big Data y la publicidad personalizada en redes sociales*. <http://uvadoc.uva.es/handle/10324/40122>



- Chicano Tejada, E. (2017). *Small data vs. big data. ¿Cuál es la mejor alternativa? ¿En qué se diferencian?* Revista Digital INESEM. <https://revistadigital.inesem.es/gestion-empresarial/small-data/>
- Cid Robles, C. (2019). *Comparación bases de datos relacionales y no relacionales*. Medium. <https://medium.com/200-response/comparación-bases-de-datos-relacionales-y-no-relacionales-424dd83dd693>
- Conesa Caralt, J., & Curto Díaz, J. (2012). *Introducción al Business Intelligence*. En *Editorial UOC (2ª)*.
- Costalago Serrano, R. (2019). *Aplicaciones de la inteligencia artificial en la industria publicitaria*. <http://uvadoc.uva.es/handle/10324/36896>
- Del Val Román, J. L. (2016). *CONFERENCIA DE DIRECTORES Y DECANOS DE INGENIERÍA INFORMÁTICA Industria 4.0: la transformación digital de la industria*.
- Espinosa, R. (2016). *Indicadores de gestión: ¿Qué es un KPI?* <https://robertoepinosa.es/2016/09/08/indicadores-de-gestion-que-es-kpi>
- Facultad de Estudios Estadísticos UCM. (s. f.). *¿QUÉ ES BIG DATA?* Máster Big Data y Business Analytics de la Universidad Complutense de Madrid. Recuperado 22 de mayo de 2020, de <https://www.masterbigdataucm.com/que-es-big-data/>
- Ferrer-Sapena, A., & Sánchez-Pérez, E. A. (2013). *Open data, big data: ¿hacia dónde nos dirigimos?* *eprints.rclis.org*, 7, 150-156. <http://eprints.rclis.org/21006/>
- Feval Institución Ferial de Extremadura. (2018). *Agricultura de Precisión con Imágenes de Satélite, Sensorización y Big Data*. Si No Se Cancela. <https://www.sinosecancela.com/extremadura/events/650430/agricultura-de-precision-con-imagenes-de-satelite-sensorizacion-y-big-data>
- Flores Vivar, J. (2018). *Algoritmos, aplicaciones y big data, nuevos paradigmas en el proceso de comunicación y de enseñanza-aprendizaje del periodismo de datos*. *scielo.org.pe*. http://www.scielo.org.pe/scielo.php?pid=S1684-09332018000200013&script=sci_arttext
- Gómez-Ruano, M. Á. (2017). *La importancia del análisis notacional como tópico emergente en Ciencias del deporte*. [The importance of performance analysis as an emergent research topic in sport sciences]. *RICYDE. Revista Internacional de Ciencias del Deporte*. doi:10.5232/ricyde, 13(47), 1-4. <https://www.cafyd.com/REVISTA/ojs/index.php/ricyde/article/view/1148>
- González-Iglesias González, R. (2015). *Herramientas para el BigData y Machine Learning*. <http://uvadoc.uva.es/handle/10324/14154>
- González Gómez, S., & Rubio Gil, Á. (2019). *Big Data y Turismo Deportivo: Estado de la cuestión y nuevas aplicaciones*. *camerablu.unina.it*, 2. <http://www.camerablu.unina.it/index.php/eracle/article/view/6194>
- González Rodríguez, F. (2020). *Sandbox, aplicación del Big Data en el futuro del Fútbol*.



- Grupo IGN. (2017). *Historia del Big Data*. Grupo IGN. <https://ignsl.es/historia-del-big-data/>
- Gutiérrez del Campo, A. (2019). *Análisis Real Time del mercado de valores con tecnologías Big Data*. <http://uvadoc.uva.es/handle/10324/38704>
- Hereña, A. P. (2018). Análisis de los ataques que terminan en gol en la Liga Santander= Analysis of the offensive plays that end in a goal in the Santander League. En *buleria.unileon.es*. <https://buleria.unileon.es/handle/10612/10949>
- ITELLIGENT. (2019). *Small Data, qué es y en qué se diferencia del Big Data*. ITELLIGENT. <https://itelligent.es/es/small-data/>
- Leguizamón Rojas, G. A. (2018). *Análisis de datos y modelos de aprendizaje para monitorizar el consumo de agua en redes de abastecimiento usando tecnologías Big Data*. <http://uvadoc.uva.es/handle/10324/33328>
- Malvicino, F., & Yoguel, G. (2016). BIG DATA. AVANCES RECIENTES A NIVEL INTERNACIONAL Y PERSPECTIVAS PARA EL DESARROLLO LOCAL. En *ciecti.org.ar*. www.ciecti.org.ar
- Martín Peña, L. M. (2016). *Aplicación de Business intelligence y Big Data*. <http://uvadoc.uva.es/handle/10324/19681>
- Martínez, E., & Ventura, J. (2016). *Árboles de decisión*. Extracción y Recuperación de Información. <https://sites.google.com/site/extraccionyrecuperaciondeinfo/home/arbolesde-decision>
- Matesanz Niño, M. (2018). *Análisis de sentimiento en Twitter con herramientas de Big Data*. <http://uvadoc.uva.es/handle/10324/33323>
- Menasalvas, Ernestina; Gonzalo, Consuelo; Rodríguez-González, A. (s. f.). BIG DATA EN SALUD: RETOS Y OPORTUNIDADES. En *minetad.gob.es*. Recuperado 26 de junio de 2020, de <http://www.minetad.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/405/MENASALVAS, GONZALO Y RODRÍGUEZ.pdf>
- Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. *sedici.unlp.edu.ar*, 278-281. <http://sedici.unlp.edu.ar/handle/10915/20034>
- Monedero Carreras, Á. J. (2019). *Análisis y visualización Big Data en eventos deportivos*. <http://uvadoc.uva.es/handle/10324/38979>
- NIÑO, M., Technologies, A. I.-D. N., & 2015, U. (2015). Entendiendo el Big Data: antecedentes, origen y desarrollo posterior. *revistadyna.com*, 2, 8. <https://www.revistadyna.com/Articulos/Ficha.aspx?idMenu=ae91868c-6abc-4b57-803c-4a45e646d0b8&Cod=7835&CambioLogin=1>
- Parra Calderón, C. L. (2016). Big data en sanidad en España: la oportunidad de una estrategia nacional. *SciELO Espana*, 30.



http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112016000100013

- Peña, D. (2002). *ANÁLISIS DE DATOS MULTIVARIANTES*.
https://www.researchgate.net/profile/Daniel_Pena4/publication/40944325_Analisis_de_Datos_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf
- Porras Blanco, M. (2017). *KPI's ¿Qué son, para qué sirven y por qué y cómo utilizarlos?* LOGICALIS. Architects of Change.
<https://blog.es.logicalis.com/analytics/kpis-que-son-para-que-sirven-y-por-que-y-como-utilizarlos>
- PowerData. (s. f.). *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*. PowerData. Recuperado 22 de mayo de 2020, de <https://www.powerdata.es/big-data>
- QuestionPro. (2018). *Análisis de Datos*. QuestionPro.
<https://www.questionpro.com/es/analisis-de-datos.html>
- Rivas González, C. (2019). *Big Data y su influencia en el consumo en la empresa privada*. <http://uvadoc.uva.es/handle/10324/38715>
- Rivera Resina, F. J. (2018). *Aplicación de Business Intelligence en una pequeña empresa mediante el uso de Power Bi*.
<http://uvadoc.uva.es/handle/10324/32877>
- Simmons, R., & Berri, D. (2019). Los mercados de trabajo en el deporte de equipo profesional. *gepacv.org*, 148-164. www.funcas.es
- Stoyanov Georgiev, K. (2019). *Implementación de un sistema de Business Intelligence en una corporación industrial*.
<http://uvadoc.uva.es/handle/10324/37872>
- tecnologias-informacion.com. (2018). *Análisis de Datos*. TECNOLOGÍAS INFORMACIÓN. <https://www.tecnologias-informacion.com/analisis.html>
- TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN II (TIC'S II). (s. f.). *MODELOS RELACIONALES DE BASE DE DATOS*. TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN II (TIC'S II). Recuperado 9 de julio de 2020, de <https://finanzastics2.wordpress.com/3-modelos-relacionales-de-base-de-datos/>
- Trabado, M. Á. (2018). *Un minuto en Internet*. <https://miguelangeltrabado.es/un-minuto-internet/>
- Wikipedia. (2020a). *Análisis de datos*. Wikipedia. La Enciclopedia Libre.
https://es.wikipedia.org/w/index.php?title=Análisis_de_datos&oldid=125538114
- Wikipedia. (2020b). *Datos Abiertos*. Wikipedia. La Enciclopedia Libre.
https://es.wikipedia.org/w/index.php?title=Datos_abiertos&oldid=125073255



Wikipedia. (2020c). *Extract, transform and load*. Wikipedia. La Enciclopedia Libre.
https://es.wikipedia.org/w/index.php?title=Extract,_transform_and_load&oldid=123511623

Zanoni, L. (2014). *Futuro inteligente*.
<https://books.google.es/books?hl=es&lr=&id=pjSiBQAAQBAJ&oi=fnd&pg=PT5&dq=el+futuro+del+big+data&ots=DE5LSR5WFe&sig=noKqrjgygVJ9RaeVNFRnWf6SZG4>

