# 2. Introducción al análisis métrico de datos

BARÇA INNOVATION HUB
Universitas


- 2.1 Analytics: el enfoque de Alamar
- 2.2 Bases de datos
- 2.3 Criterios de selección de datos
- 2.4 Uso y transferencia de datos
- Referencias

# 2.1 Analytics: el enfoque de Alamar

El término *analytics* (analítica) ha expandido su uso en el ámbito deportivo en la década de 2010. Como indican Jayal, McRobert, Oatley (2018) no es otra forma de nombrar a lo que previamente se denominaba análisis, sino que implica una vinculación a la toma de decisiones en deportistas, entrenadores y gestores.

Benjamin Alamar (2013) establece conclusiones a partir de una encuesta de uso de *analytics* en instituciones deportivas (Encuesta "SAUS") sobre el impacto en los procesos de tomas de decisiones.

**Figura 1: Benjamin Alamar**

Fuente: [Imagen sin título sobre Benjamin Alamar]. (s/f). Recuperada de https://www.linkedin.com/in/benalamar?trk=public_profile_browsemap_mini-profile_title.

Esta encuesta “SAUS” se desarrolló en clubes NFL (fútbol americano), MLB (béisbol), NBA (baloncesto) y Premier League (fútbol inglés). Entre las principales conclusiones, Alamar indica que no suelen existir visiones semejantes dentro del mismo club respecto a la percepción del grado de calidad alcanzado por los sistemas de análisis. Es habitual que dos ejecutivos del mismo club tengan visiones dispares sobre el nivel de adecuación y aprovechamiento de una base de datos o de la aplicación usada en el club. Es decir, en los clubes profesionales no suele existir una visión compartida internamente sobre cuál es el estado de ventaja competitiva respecto de los datos obtenido en comparación a los rivales.

La película “Moneyball”, basada en la metodología empleada en béisbol desde la década de 2000 por los A’s de Oakland, permitió una menor resistencia por parte de público, prensa y gestores al uso de datos en los deportes. Paul DePodesta fue el principal asistente de datos del ejecutivo Billy Beane en los A’s.

**Figura 2: Paul DePodesta**

Fuente: [Imagen sin título sobre Paul DePodesta]. (2017). Recuperada de https://commons.wikimedia.org/wiki/File:Paul_DePodesta_(36380255290).jpg.

Como ejemplo de la posible transferencia de **analytics** a otras disciplinas deportivas, se encuentra el hecho de la transición de Paul DePodesta desde el béisbol al fútbol americano, en el que actualmente es director de estrategia de los Cleveland Browns en la NFL.

El enfoque estadounidense en datos deportivos ha sido más industrial que el europeo, el cual se ha centrado en la investigación, la comprensión del juego y una perspectiva a largo plazo, a pesar de autores como Peter O’Donoghue (profesor de la Universidad Metropolitana de Cardiff) que ha vinculado sus obras a las tomas de decisiones. Los centros de alto rendimiento y de tecnificación europeos suelen disponer de laboratorios de investigación que permiten un conocimiento más asentado y estable, pero de mayor dificultad en su

aplicación directa e inmediata a las ventanas de mercado de adquisición de jugadores o a los planes de partido.

### 2.1.1 *Whitepaper*: la visión de *Davenport*

Thomas Davenport, un experto en *analytics* procedente del mundo de la empresa identificó el deporte como el nuevo campo más prometedor para la aplicación de datos, por lo que realizó un proyecto de 25 entrevistas a expertos pertenecientes a equipos profesionales de fútbol europeos y del deporte estadounidense, no solo entrenadores y ejecutivos de estos clubes, sino también expertos en sistemas informáticos.

Unos de los principales factores de éxito identificados por Davenport (2014a) fue el acceso a vídeo de forma limpia y rápida. No depender del intercambio de vídeos con otros equipos mediante la disposición de plataformas corporativas de ligas puede ser un pilar sobre el que se construya la cultura del dato, tanto para el registro customizado como para la visualización indexada mediante metadatos.

Entre sus conclusiones, Davenport (2014a) distingue en 3 áreas diferentes la toma de decisiones basada en datos:

**Rendimiento deportivo**

- En el que identifica como principales factores diferenciales el uso de datos posicionales (*tracking*), la implicación e involucración de jugadores en la visualización de informes o la creación de bases de datos propias con variables customizadas.

**Gestión deportiva**

- El foco principal es el estudio del *fan* o asistente al estadio o instalación deportiva, mediante redes Wifi que permitan estudiar sus desplazamientos y consumos, así como segmentaciones de precios de entradas o acciones de *marketing*

**Salud**

- La clasificación de lesiones y su posible predicción es prioritaria.

Algunos de los factores que Davenport (2014a) asocia a buenas prácticas en cuanto a análisis métricos son:

- Alineamiento y visión común entre responsables de tomas de decisiones.
- Las ligas actúan como responsables de la coordinación de inversiones.
- Los clubes compran y adquieren datos que pasan a pertenecerles.
- Búsqueda de los clubes de asesoramiento por parte de consultores y de la industria, estableciendo alianzas.
- Establecer proyectos que permitan a voluntarios participar en el club mediante visiones alternativas en el tratamiento de datos.
- Crear índices combinados que unan variables deportivas y financieras.
- Explorar las posibilidades de análisis predictivos y prescriptivos.

## 2.1.2 *Game analysis*: la adaptación de O'Donoghue

Peter O'Donoghue realiza una revisión (2014) de distintas soluciones de software disponibles para el análisis del *play-by-play* (jugada a jugada) que permiten profundizar en el conocimiento del juego, como *FocusX2*, *SportsCode*, *Dartfish*, Excel, SPSS o MATLAB. También mostró las posibilidades de datos posicionales de fútbol con coordenadas X, Y de jugadores y balón proporcionados por Prozone (actualmente STATS). Su adaptación al game analysis puede estudiarse en las publicaciones periódicas de la revista académica *International Journal of Performance Analysis in Sports,* de la que es editor. O'Donoghue (2014) indica escenarios en los que el registro de datos se realiza manualmente por un observador y también en los que se realiza una importación desde proveedores de datos.

Lebed (2017) muestra la transición progresiva que se produce al transformar los datos en conocimiento, incidiendo en un paso posterior, que es el de wisdom (sabiduría), necesario para la toma de decisiones en deporte.

## Anexo I

Supuesto práctico respecto a la introducción al análisis métrico de datos.

En junio de 2015, la revista *Bloomberg Businesswek* publicó un monográfico titulado "*What is code?*" ("¿Qué es el Código?") sobre lenguajes de programación informática, código y desarrollo, un ensayo escrito por el autor Paul Ford (2015). En dicho número, se intercalaron continuamente dos bloques de contenido: uno de información y otro de ficción sobre la posible aplicación de esa información en la industria del software. Tomamos esa idea para recrear un escenario posible, práctico y ficticio de la industria deportiva y el análisis métrico de datos.

Antonio espera en la sala su turno para acceder a la entrevista de trabajo a la que ha sido convocado. Es 2011, las tablets aún son dispositivos lentos que no se han expandido, por eso ha preparado algunas notas en su *BlackBerry*, que todavía compite con Android en el mercado. Ha impreso algunas gráficas y tablas que porta en una carpeta con separadores de plástico, por si fuera oportuno presentarlas en la entrevista.

La agenda de la *Blackberry* acumula varios avisos que informan sobre el retraso de los entrevistadores. Por fin, es su turno. Se trata de una empresa que gestiona la entonces vigente "Ley Orgánica de Protección de

Datos" de un club de la élite de los deportes de sala que no cuenta con secciones de fútbol, él escribió por correo electrónico y llamó espontáneamente hace unas semanas para proponer un proyecto de investigación en gestión de datos ligado a la tesis doctoral que ha iniciado recientemente en la universidad de la misma ciudad del club.

Tras cinco minutos de conversación, la decisión está tomada y comunicada: Antonio trabajará para la empresa de datos en el club desde el siguiente día. No queda muy definido su rol ni su sueldo, le quieren tener ahí "para que haga algo con los datos". Colaborará con el equipo filial en funciones de registro de *tests* físicos, lesiones, etc., aunque él, por el momento, no comenta nada sobre lo que le apasiona, el análisis de competición.

Es el primer entrenamiento, y Antonio es el primero en aparecer con un traje de oficina en el ambiente de la pista, balones, entrenadores y jugadores con camiseta y pantalones cortos. Le encomiendan sentarse en una silla llena de resina que le destroza el traje y tras presentarse, el delegado de la empresa en el club, le encomienda realizar un registro manual sobre una planilla fotocopiada en la que debe anotar la asistencia de jugadores al entrenamiento.

El segundo día recibe ropa deportiva del club con el anagrama y sus iniciales serigrafiadas (ha caído bien en el filial tras conversaciones con el cuerpo técnico en el entrenamiento anterior). Tiene un Excel abierto en la *Blackberry* que envía por correo electrónico al segundo entrenador del equipo al finalizar el entrenamiento, en el que aparecen las asistencias y ausencias a entrenamientos, que progresivamente sustituye a la fotocopia que recibió.

Está todo por hacer, pero no recibe ninguna misión: simplemente, hay que hacer algo con los datos, generarlos, almacenarlos… y obtener alguna conclusión sobre ellos. Su día a día se separa cada vez más de los objetivos de la tesis doctoral que finalmente abandona (mejor dicho, no la inicia) para centrarse en el club.

Recuerda el nombre de una revista estadounidense deportiva: "*Insight*". Eso es lo que quiere el club de él, y también la empresa. Generar *insight*, una visión profunda que genere un impacto. El cuerpo técnico del primer equipo empieza a interesarse en los registros individuales de minutos de entrenamientos y partidos que lleva a cabo para el filial y empiezan a proponerle registros y a conversar con él. Unos días antes de finalizar la pretemporada, recibe la invitación del preparador físico del primer equipo para colaborar en la realización de un test.

Uno de los jugadores llega tarde al entrenamiento de los tests y el cuerpo técnico comienza a sonreír cuando observa a Antonio registrar en la *Blackberry* una "semi-falta" de asistencia. Empieza el entrenamiento del filial en la pista contigua, y Antonio es saludado con un afable "ya no quieres nada con nosotros" por parte del cuerpo técnico del filial, al que responden los técnicos del primer equipo con "este lo hemos subido igual que el pivot el verano pasado, que empezó la pretemporada con vosotros".

Tras tabular y mostrar los datos al preparador físico, Antonio recibe una indicación para crear grupos de entrenamiento basados en los valores de resistencia observados. Tiene una idea que es aceptada: además de generar 3 grupos propuestos por el preparador, se puede realizar un código de colores de semáforo en el que el verde indique que se está por encima del percentil 75, recogido en la bibliografía científica en ese deporte y el rojo por debajo del percentil 25. La presentación se imprime en papel tras cada test, hasta que, a mitad de temporada, 5 meses después, se crea una omisión de Evaluación en la que dispondrá de 5 minutos para mostrar al cuerpo técnico, ejecutivos y directivos algunos de los resultados de su trabajo. Será la primera vez que le solicitan realizar una presentación en PowerPoint.

El viejo proyector del club juega una mala pasada y no discrimina rojos de verdes, igualándolos en un ocre que dificulta la comprensión de los datos presentados. Sin embargo, el cuerpo técnico no esperaba tanta transparencia en una reunión en la que sienten que su trabajo es juzgado negativamente, entre otros aspectos, por los datos de resistencia mostrados, tanto de tests como de tracking. Antonio no consultó previamente al preparador físico si debía mostrar los datos de aquel test y se arrepiente de no haberle solicitado permiso.

Desde aquella tarde, Antonio pierde la confianza de los técnicos, pero los directivos solicitan durante el verano su incorporación directa al club, que se lleva a cabo al iniciarse la pretemporada. Tampoco recibe ninguna misión concreta: su labor parece ser la de un analista de datos "a demanda" que debe generar informes… a cualquier directivo que se lo proponga.

Tras una semana de incorporación al club, Antonio ha recibido solicitudes de informes del departamento financiero, presidencia, gerencia, médico, etc., sobre datos deportivos que pueden ser vinculados a contratos de jugadores para calcular su eficiencia o para determinar la idoneidad de cambios en procesos de gestión. Sin embargo, no recibe ningún encargo del área deportiva, a pesar de trabajar con datos: ha roto una de las "leyes no escritas": ha contado fuera del vestuario lo que se hace dentro, aunque sea en forma de datos de un test.

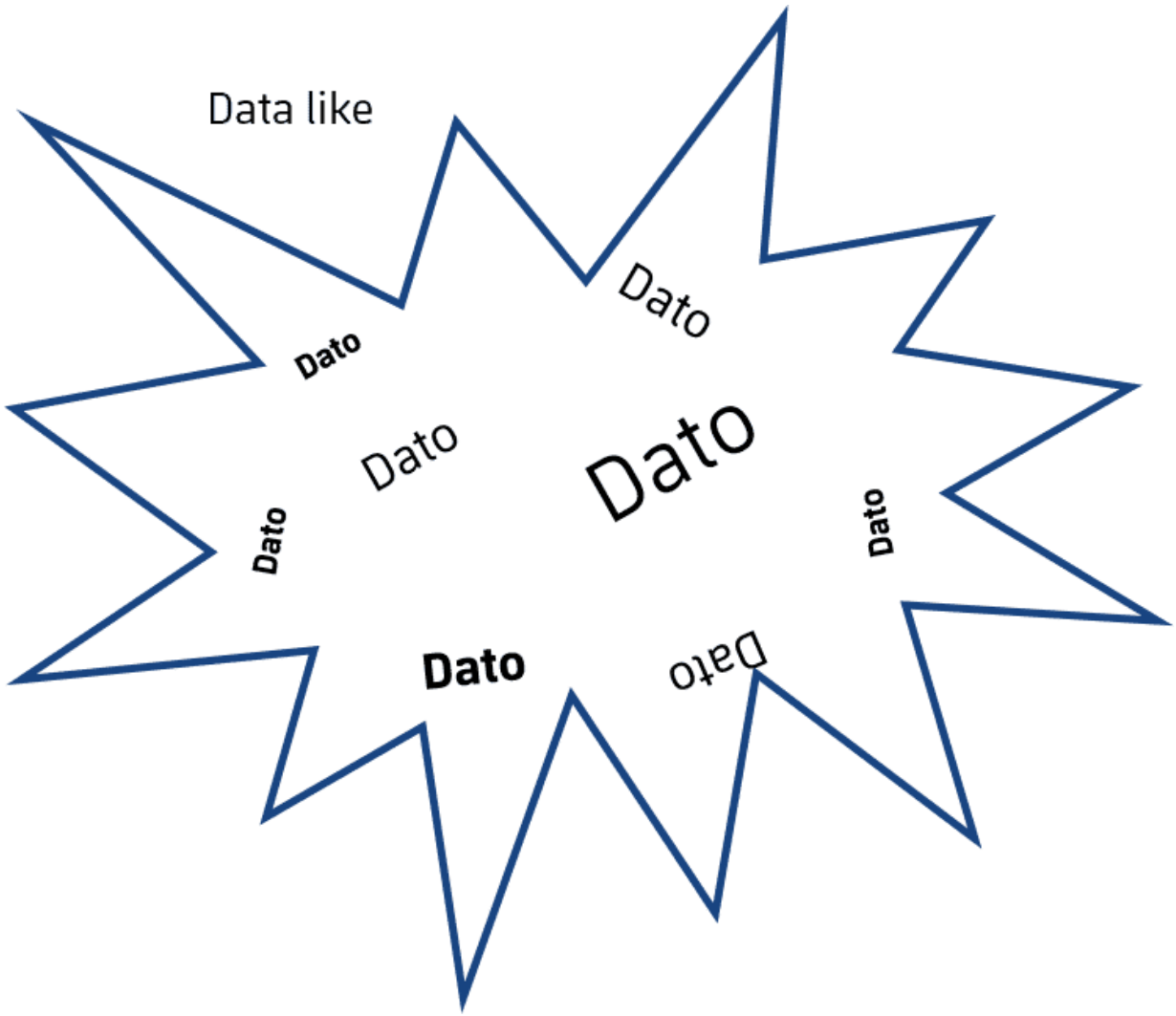
CONTINUAR

# 2.2 Bases de datos

## 2.2.1 Almacenamiento: *data lake*

Una vez adquiridos o registrados los datos deportivos, el siguiente paso es establecer un repositorio único en el que sean almacenados digitalmente: este es el concepto de data lake.

**Figura 3: Data lake**




Fuente: Elaboración propia.

El *data lake* puede disponerse en un servidor propio del club, o mediante almacenamiento acceso *cloud* (en la nube), con ventajas e inconvenientes en cada caso.

Si se opta por un *data lake* físico, se dificulta el acceso a vídeos y archivos importantes cuando se disputa un partido en otra ciudad o se intenta acceder desde el domicilio de los miembros del cuerpo técnico, aunque es posible (pero de forma más lenta) con soluciones de tipo VPN (red privada virtual).

Si la solución elegida es un *data lake cloud,* se abaratan los costes de mantenimiento y alojamiento para el mismo volumen de información, además de permitir un acceso desde cualquier ubicación. Como inconveniente, la velocidad de acceso al dato (especialmente importante cuando se accede a archivos de

varios Gb de información, por ejemplo, vídeos o archivos de *Raw Data* de partidos o entrenamientos) encarece progresivamente el precio de *cloud*.

En cualquier caso, el principio fundamental es evitar silos de información en los que la información del club se almacene en distintos repositorios no conectados entre sí, y para esta función primordial el *data lake* es una solución adecuada.

## 2.2.2 Estructuración: *data warehouse*

Una vez centralizados y almacenados en formatos básicos la información en un *data lake,* se precisa de una labor de estructuración y jerarquización mediante un data *warehouse*.

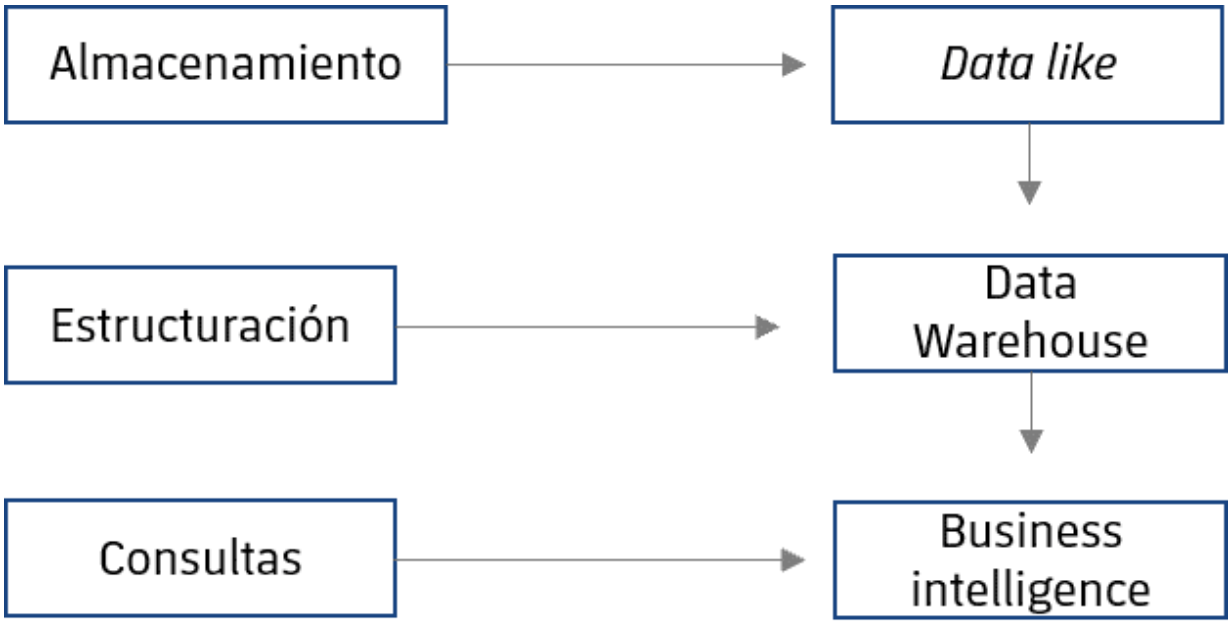
El *data lake* permitirá las funciones ETL (E=Extracción, T=Transformación, L=Load - carga -) de datos ordenados en subunidades denominadas data marts.

Para llevar a cabo las funciones de transformación en el *data lake* se necesita una labor previa de definiciones y jerarquización de datos básicos que permita su adecuada gobernanza.

## 2.2.3 Consultas: *Business Intelligence*

Tras el almacenamiento en *data lake*, la preparación y jerarquización en *data warehouse,* se "sirve el consumo de datos" mediante consultas en una solución *Business Intelligence* con aplicaciones que disponen de visualizaciones y posibilidades de filtros sin necesidad de conocimientos tecnológicos avanzados.

**Figura 4: Proceso de utilización de base de datos**




Fuente: Elaboración propia.

## Anexo II

## Supuesto práctico de bases de datos

Es el inicio de la tercera temporada de Antonio en el club y existe un mantra que se escucha en pasillos y oficinas del club con: "ERP", "ERP", "ERP"... existe entre los directivos del club una antigua preocupación por la pérdida de información cuando un ejecutivo o cuerpo técnico abandona el club y se lleva con él, y su ordenador, información sensible en hojas de cálculo y aplicaciones no centralizadas. Se empieza a hablar hace unos meses de la necesidad de una **base de datos centralizada**, en las primeras reuniones para abordar este objetivo el departamento financiero explica al deportivo que pueden usar la solución de la que ellos disponen porque "es modular". A las primeras preguntas a los proveedores informáticos sobre cuánto tiempo tardaría en estar disponible un registro de ejercicios de entrenamiento siguen silencios que ponen en duda la idoneidad de esas soluciones.

Antonio recibe el encargo de realizar una prospección de mercado. Presenta varias soluciones con un análisis DAFO y precios aproximados en una reunión posterior, a la que el club termina inclinándose por la solución planteada por el "Informático" (en la que aclara que él no es informático, sino que es administrativo con conocimientos de informática que trabaja en ese departamento) del club. Esa solución fue previamente evaluada por Antonio, en la que indicaba que se obtendrían como fortaleza *quick wins* ("éxitos rápidos") y

como amenaza, que la solución elegida alcanzaría también de forma rápida un techo infranqueable al no disponer actualmente de importaciones o exportaciones o posibilidad de comunicación con otras bases de datos del club. La principal debilidad es que no se cumpliría el objetivo de centralizar todos los datos generados en el club.

Pocas semanas después, Antonio recibe tres buenas noticias:

- Se aprueban convenios de prácticas con universidades que le permitirán disponer de alumnos de distintas titulaciones a los que tendrá que formar en el ETL de datos, pero de los que también podrá obtener informes de lesiones y de rendimiento físico.

- Se aprueba la incorporación de un asistente que le apoye en su trabajo diario.

- La Federación dotará, en la próxima temporada, una nueva aplicación informática que proporcionará (en tiempo real y post-partido) múltiples datos de toda la competición de los que ahora no disponen.

El ritmo de trabajo aumenta drásticamente en las oficinas. Alumnos en prácticas, vestidos con la ropa cedida temporalmente durante los periodos de dos meses que dura su estancia, recorren con tablets corporativas las pistas de entrenamiento, recogen la percepción subjetiva del esfuerzo de los jugadores y las centralizan en la nueva aplicación comprada. El nuevo asistente se encarga de las coordinaciones rutinarias en el Centro de Tecnificación cuando Antonio debe acudir, cada vez más, a la capital donde tiene la sede la liga para reuniones temáticas y sectoriales en las que representa al club.

Una reunión del Consejo de Administración, que se realiza en una de las noches en las que regresa de las reuniones en la capital, tiene como inesperada consecuencia la caída de la cúpula de la Junta Directiva, produciéndose una renovación casi total de los órganos de dirección y ejecutivos en la nueva estructura del club.

Ante este nuevo panorama, Antonio presenta en la reunión mensual de Directores su *briefing*. Muestra con entusiasmo las dos nuevas aplicaciones informáticas que la Federación ha creado, y que pueden brindar una ventaja competitiva para el club por el trabajo desarrollado previamente.

La cultura del club ha cambiado: Antonio informa que deben adquirirse al menos dos nuevos dispositivos para ejecutar las nuevas aplicaciones, un iPad para el *software del live* de partidos y un ordenador Windows potente para los post-partidos. El club no dispone actualmente de ninguno de los dos dispositivos, y las aplicaciones deben ser pagadas obligatoriamente por todos los clubes pertenecientes a la máxima competición de la Federación.

Comienza la temporada y ni el iPad ni el ordenador Windows "potente" han llegado. Están estudiando el gasto, le dicen durante la última semana cuando acude varias veces al despacho del gerente. En el primer partido de liga, Antonio ha pedido prestado a un médico su iPad para instalar la aplicación de live

Es lunes, y Antonio recibe una llamada de la Federación preguntándole por qué no ha iniciado la clave asociada a la IP que activa la nueva aplicación post-partido. Explica que el club todavía no ha dotado a su departamento del ordenador con los requisitos necesarios. Desde la Federación le comentan que no es el único caso.

Tras unos meses, se recibe el ordenador: es la jornada 16.

CONTINUE

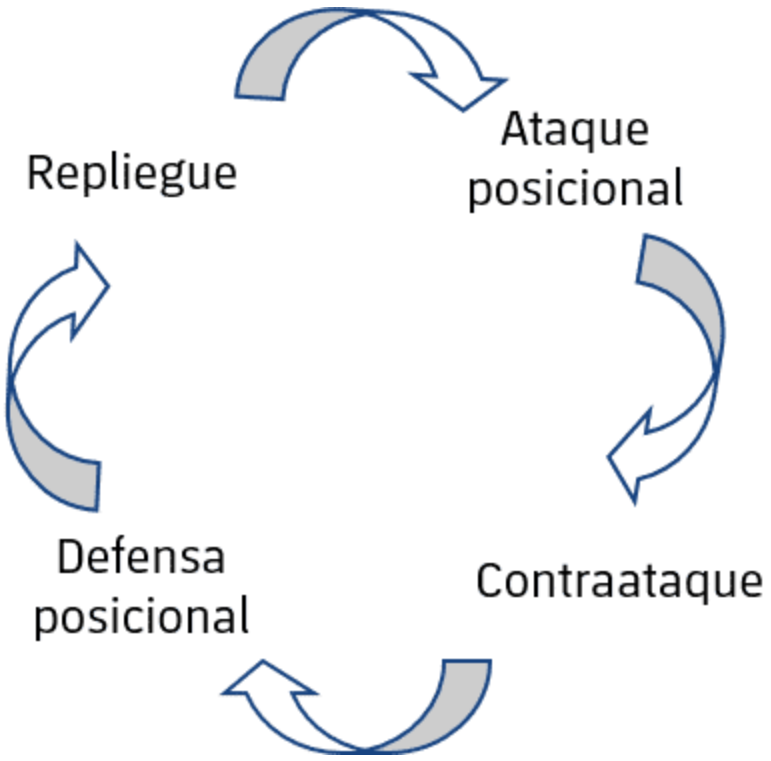
# 2.3 Criterios de selección de datos

## 2.3.1 *Whitehead*: saber observar es saber seleccionar

Una adecuada selección de variables a observar suele venir determinada por proveedores externos de datos que utilizan una selección lo más amplia posible para abarcar distintas perspectivas y necesidades de diferentes clientes, en las que suele estar ausente una estructuración de las fases y modelos de juego.

Un primer punto de partida para llevar a cabo la selección de métricas a incluir en un modelo de análisis, es establecer las fases del modelo deportivo en cuestión. Por ejemplo, en deportes colectivos es habitual el modelo de 4 fases.

**Figura 5: Modelo de juego del deporte**



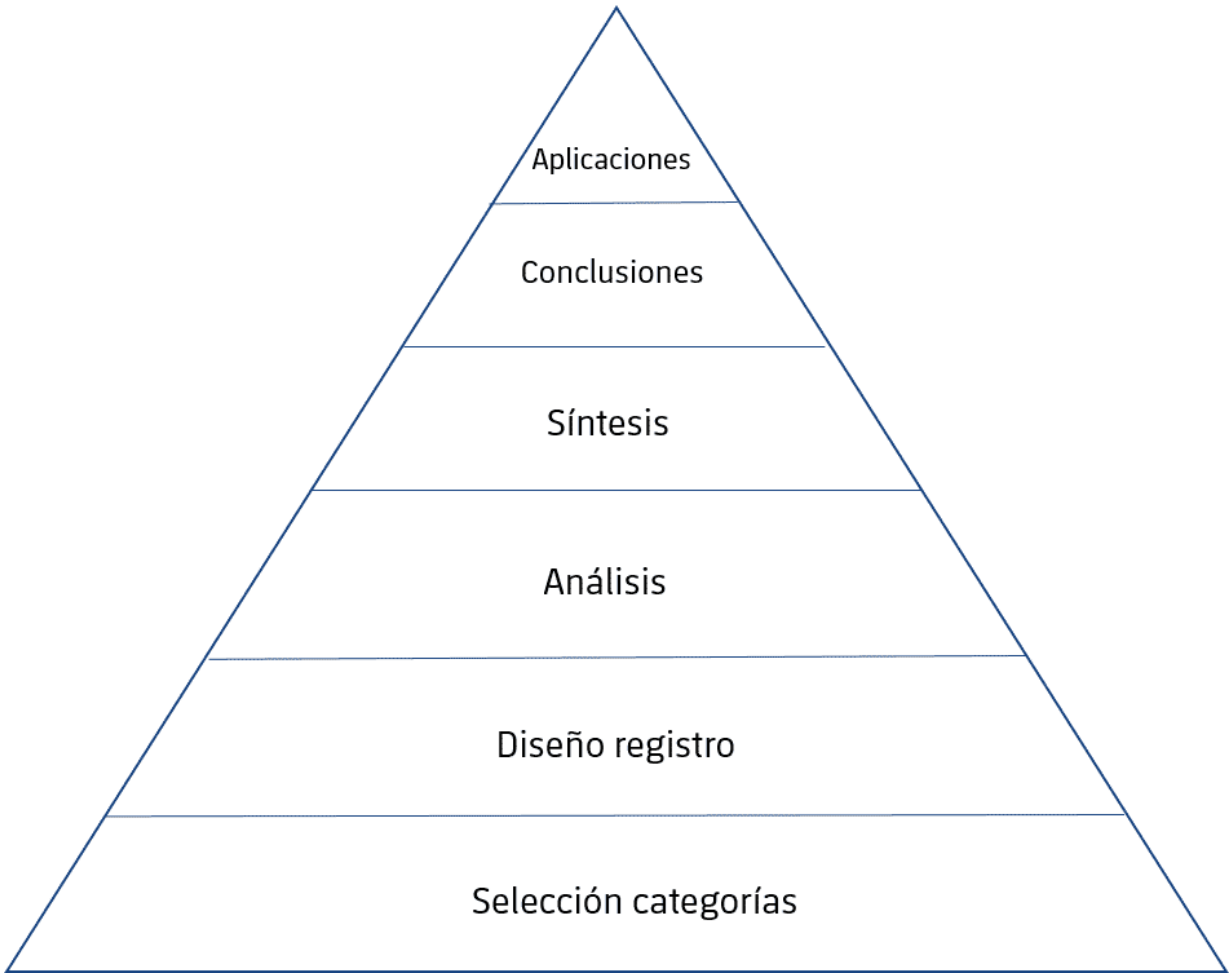
Fuente: Elaboración propia.

Indicó Whitehead (1967) que “saber observar es saber seleccionar” (p. 28). Todos los modelos son aproximaciones que omiten las partes menos relevantes, pero que permiten una comprensión de un fenómeno. Existen dos criterios fundamentales para realizar la selección de variables a observar:

- El cuantitativo, en cuanto al impacto de la variable en el resultado final.
- El cualitativo, referido a las relaciones claves en el encadenamiento de acciones, aunque la variable examinada no tenga un alto impacto por sí sola en el impacto final.

Tras la selección de categorías, basada en un adecuado modelo de de juego son necesarios:

**Figura 6: Pirámide de utilización de base de datos**




Fuente: Elaboración propia.

**Diseño registro**

No son casos similares el de una app de registro en directo de partidos para suministrar información durante el descanso, que una de análisis post-partido, en la que se puede disponer de mayor número de operadores y tiempo. Existen limitaciones de personal y temporales que pueden determinar incluso el paso anterior de selección de categorías.

**Análisis**

Una vez disponible el instrumento de registro ajustado a las categorías seleccionadas, se inicia la observación en sí con una digitalización adecuada.

**Síntesis**

Poder filtrar y reajustar de manera abierta cualquier tipo de consulta no es posible en muchas apps y soluciones de *software* que *preajustan* el tipo de reporte accesible. Poder acceder a cualquier tipo de filtro y combinación de variables registradas es dar sentido al esfuerzo realizado previamente en el análisis y registro.

**Conclusiones**

La transformación de datos en conocimiento da lugar a conclusiones descriptivas no prescriptivas a partir de informes.

**Aplicaciones**

El impacto en el entrenamiento o la competición del análisis métrico de datos, puede darse a partir de la interpretación del cuerpo técnico de las conclusiones previas.

## 2.3.2 Los 5 tipos de datos deportivos

Cualquier tipo de dato presente en la industria deportiva puede clasificarse dentro de los siguientes 5 bloques:

- *Eventing*.
- *Tracking*.
- Biológicos.
- Financieros.
- Vídeo.

Los datos de *eventing* son aquellos que indican secuencias de acciones técnico-tácticas registrables, como tiros, pases o regates en fútbol, asociados a una marca temporal. También se incluyen aquí datos como la fecha y lugar del evento deportivo o los nombres de deportistas.

En cuanto al *tracking* o datos posicionales, son las coordenadas espaciales ocupadas en la instalación deportiva por atletas, jugadores, implementos u objetos en determinadas frecuencias (una frecuencia de muestreo de 1 Hz indicaría una toma de datos por segundo, 50 Hz indicaría una muestra cada 0,02 segundos).

Los datos biológicos son todas aquellas analíticas biomecánicas o fisiológicas (sanguíneas, mediciones de peso y composición corporal, lesiones, frecuencia cardiaca, producción de fuerza) de deportistas. Este tipo de datos también se conoce como **vital** en el ámbito centroeuropeo.

Los datos financieros, más usados en el contexto de la industria deportiva estadounidense, se ocupan de cantidades económicas referidas a sueldos, fichas, traspasos, etc.

El hecho de considerar al video como dato es un cambio de enfoque que se produjo con su digitalización, pero que no ha calado aún en los hábitos de la mayoría de instituciones deportivas europeas, lo que puede suponer una ventaja competitiva para los clubes que desarrollen este concepto. Vincular en cualquier base de datos el video a un filtrado de variables puede ser el paso más omitido para transmitir a jugadores la información de forma eficiente.

Existe otra clasificación transversal de datos que influye en el tipo de conclusiones y aplicaciones a obtener. Hohman, Lames y Letzelter (2005) consideran el entrenamiento como una **preparación de la competición.** Si se amplía este enfoque, la competición en el periodo de formación deportiva en jóvenes jugadores es un tipo de competición diferente al alto rendimiento, en la que deben aplicarse otras perspectivas en el uso de los datos.

**Figura 7: Enfoque transversal**

Fuente: Elaboración propia.

## 2.3.3 Estandarización y customización

La estandarización de los 5 tipos de datos referidos previamente ha sido habitualmente llevada a cabo por empresas privadas que fueron las primeras en dotar a un nicho de mercado o las que alcanzaron el liderazgo respecto a la competencia y generaron un efecto de imitación o copia por parte de otros *stakeholders* de la industria deportiva.

**Tabla 1: Ejemplos de los 5 tipos de datos deportivos**

| Ejemplos: estándares de datos deportivos | |
|---|---|
| Eventing | F24 (Opta) |
| Tracking. | EPTS (FIFA) |
| Biológicos | OSICS |

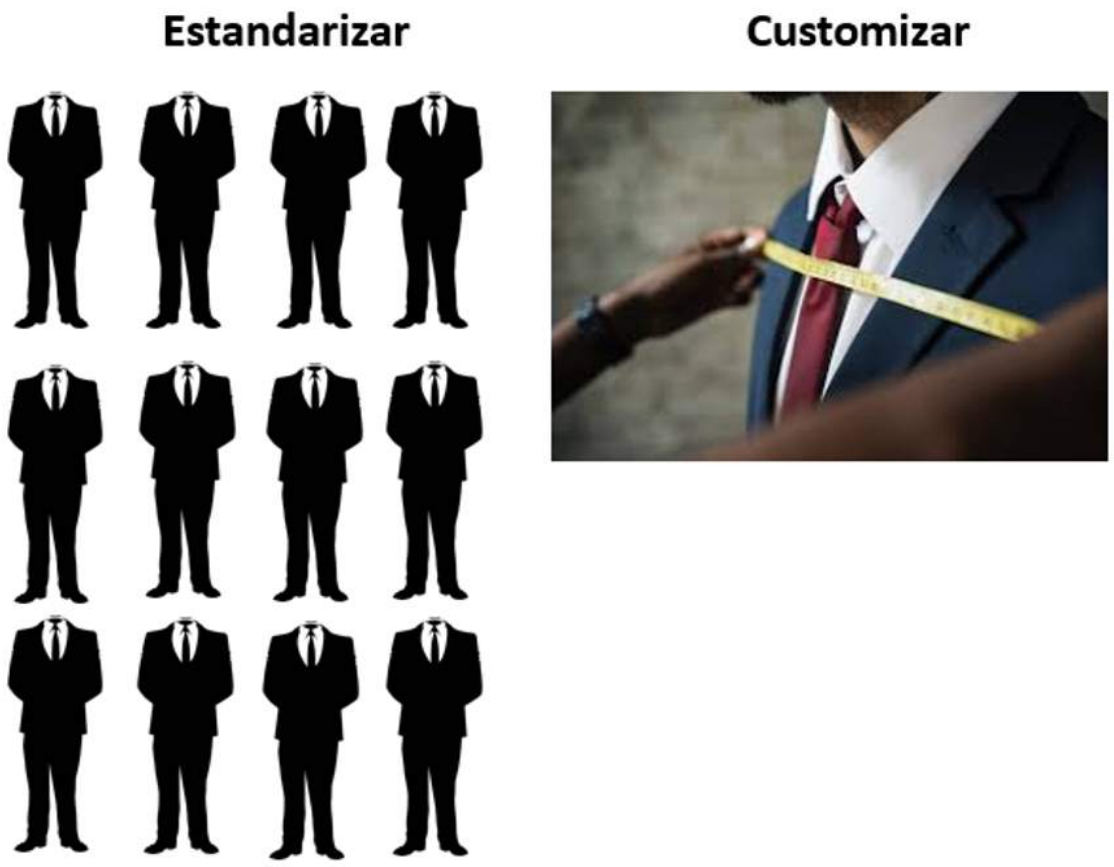
| | |
|---|---|
| Financieros | *Fair play* financiero (UEFA) |
| Vídeos | PAL, DCI 4k,<br><br>H.264/MPEG-4 AVC |

Fuente: Elaboración propia.

En los últimos años, instituciones como FIFA y UEFA han aportado estándares que pueden indicar una nueva senda en el liderazgo de datos deportivos.

En todo caso, la *customización* o personalización es necesaria para obtener ventaja competitiva en deporte.

**Figura 8: Diferencia gráfica entre estandarizar y customizar**



Fuente: Elaboración propia.

La clave reside en cuál es el punto de partida de la *customización*: empezar de cero puede ser no eficiente, en ocasiones se obtiene un rendimiento plausible al crear nuevos índices combinados a partir de métricas aisladas en los estándares.

## Anexo III

## Supuesto práctico de criterios de selección de datos

Tras una larga travesía en el desierto, el departamento de I+D+I (Investigación + Desarrollo + Innovación) del club puede analizar el *tracking* y el *eventing* de los partidos con la app que la Federación ha dispuesto. Los alumnos en prácticas proceden a realizar cada lunes exportaciones en formato csv del *play-by-play* (jugada a jugada) que posteriormente serán tratados en un Excel para generar *insights* que no alcanzan los reportes estándares de la Federación. Las posibilidades de combinación de datos son casi ilimitadas, los recursos son suficientes, la comunicación con el nuevo cuerpo técnico es fluida: casi el único factor limitante pasa a ser el tiempo o la gestión eficiente de este.

El segundo entrenador del primer equipo consulta habitualmente en los pasillos "¿Se puede calcular...?", "¿Se puede saber...?" y genera un glosario de consultas que se incorporan a un "Manual de procesos de Métricas del primer Equipo" que intentan sistematizar los flujos más habituales de consultas, aunque siempre aparecen nuevas consultas no previstas en las horas previas al partido. *Rankings*, valores de referencia, contribuciones individuales de jugadores, etc. pasan a ser sistematizados en informes post-partido, y los reportes previos al partido se aplican teniendo en consideración el rival.

El departamento médico comienza a interactuar con datos de acelerómetros que permiten reuniones quincenales sobre cargas de entrenamientos y partidos. Es un momento dulce para los datos y su aplicación, el equipo ocupa una buena posición en la tabla que le permite jugar competición europea al final de la temporada. Pero, el principal patrocinador del club quiebra y el futuro es incierto. Las vacaciones son una sucesión de consultas a apps bancarias para comprobar si se ha recibido el importe de alguna de las nóminas atrasadas, rumores de ERE y llamadas sondeando si se está dispuesto a aceptar una reducción del 30% de sueldo. Finalmente, tras el segundo día de reincorporación al club luego de las vacaciones, Antonio

y cinco compañeros son recibidos en la oficina del director general que les comunica su despido y les da cita con el Departamento de Administración para firmar su salida.

Esta vez el tren le ha pillado en marcha: uno de los proveedores de *eventing* y *tracking* llevaba tiempo ofreciéndole incorporarse a su Departamento de I+D, aunque también tendría una labor de ventas como comercial a otros clubes. Antonio, tras visitar la sede europea, acepta. Será el responsable para la Península Ibérica de uno de los mayores proveedores de datos deportivos.

La empresa dispone ya de un buen estándar de datos, de hecho, es creadora de un formato que ha sido replicado y copiado por la competencia, que trabaja con costes inferiores y se está perdiendo una cuota de mercado en la que previamente casi copaba el 100%. Antonio debe realizar entrevistas en los clubes de la "Zona Iberia" que permitan customizar informes a medida del deseo de los cuerpos técnicos. Sus reuniones son productivas y le permiten entender la mentalidad de rivales a los que ha estudiado y se ha enfrentado durante un lustro. Pero al realizar la llamada o enviar el correo electrónico con el precio del nuevo producto, tras una "demo" de varias semanas, ninguno de los clubes accede a contratarlo.

Han sido unos meses de aprendizaje de procesos más sistemáticos en la empresa, del inglés y el portugués, de los modelos de juego de los diferentes cuerpos técnicos y cuenta con la paciencia de la empresa para obtener su primera "venta", pero tras el primer semestre es absorbida por un grupo que a su vez adquirió al principal competidor. En unas semanas, Antonio debe estudiar los métodos estandarizados de la antigua competencia, que sustituyen a la estrategia customizada que se aplicaba previamente.

Una Federación de otro continente se pone en contacto con él durante un congreso de *eventing* y *tracking*. Le proponen incorporarse a su nuevo departamento de análisis con el objetivo de implementar procesos de análisis de datos "desde cero", en un país que difícilmente se va a clasificar para un Mundial y con mayor dificultad a unos Juegos Olímpicos. Le permiten trabajar la mayoría del tiempo "en remoto" y viajar puntualmente cada dos meses. Decide cambiar de trabajo.

La Federación para la que pasa a trabajar Antonio dispone de recursos básicos para llevar a cabo su nuevo "Proyecto de implementación de toma de decisiones basados en datos", inspirado en las "organizaciones *data-driven*" que ha conocido en jornadas y *synmposiums*. Los Excels que maneja la Federación empiezan a bloquearse cuando alcanzan 7-8 Megas de tamaño al incluir a nombres y fechas de nacimientos de todos los jugadores registrados, es necesario disponer de una solución de base de datos SQL creada por la propia

Federación para lo que se oferta un puesto de trabajo que permite contratar un desarrollador que pasa a ser el principal “aliado” de Antonio.

CONTINUE

# 2.4 Uso y transferencia de datos

## 2.4.1 ¿*Big Data o Smart Data*?

Tal como indica Davenport (2014b) el término *Big Data* ha sufrido de sobreuso en el ámbito del deporte, muchas veces por presión comercial. Podemos clasificar al *Big Data* como al ámbito de datos en el cual confluyen **las tres V:** volumen, velocidad y variedad. Recientemente se ha añadido una cuarta V: veracidad. Sin embargo, no se dispone de límites consensuados a partir de los cuáles se considera *Big Data* a un grupo de datos. Suele asociarse *Big Data* también a datos no estructurados o que no siguen una estructura convencional de filas y columnas.

En mayor medida que *Big Data* (grandes datos), en deporte es interesante acceder al *Smart Data* (datos inteligentes), es decir, a las técnicas que pueden permitir el acceso y explotación del dato útil.

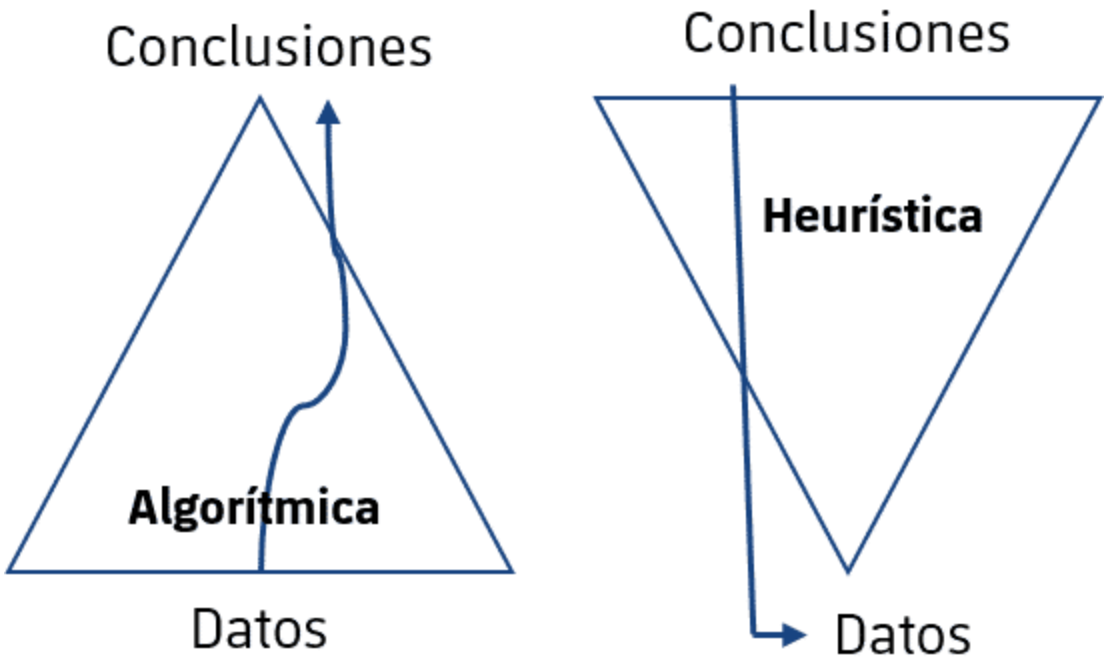
Suelen utilizarse determinados términos que pueden disociarse:

- **Inteligencia artificial:** abarca los métodos que permiten aprender de una estructura de datos para alcanzar un objetivo.
- ***Machine learning:*** es una rama de la inteligencia artificial que detecta patrones sin instrucciones previas.
- ***Deep learning:*** otra variante de la inteligencia artificial que asimila representaciones de datos.
- **Redes neuronales:** aplicadas a la resolución de problemas, permite asimilar representaciones de datos.

El enfoque más habitual por parte de entrenadores y miembros de cuerpos técnicos es heurístico-inductivo, en el que se parte de una serie de intuiciones y concepciones de la competición deportiva que pueden ser contrastados con datos.

Con una perspectiva opuesta, los responsables de datos de clubes suelen acercarse al fenómeno deportivo de forma algorítmica-deductiva en el que una de las prioridades es evitar sesgos y trasladar asociaciones de datos como conclusiones solo cuando son confirmadas de forma significativa.

**Figura 9: Diferentes enfoques de tratamiento de datos**




Fuente: Elaboración propia.

## *2.4.2 Know how:* "saber cómo" deportivo

En el proceso de madurez y mejora de un técnico deportivo, suele partirse de un periodo inicial en el que no se discrimina lo importante de lo urgente, ni lo más relevante de lo accesorio. Cuanto antes se diferencien y potencien los factores más relevantes asociados al rendimiento deportivo, en mayor medida se adecuará un método de entrenamiento y preparación de la competición.

En esa línea, una base de datos o cuadro de mandos que influya en el rendimiento debe atender, sobre todo, a aquellas situaciones del juego más relevantes, que no siempre son las más frecuentes.

**Tabla 2: Ejemplo de cuadro de mandos asociados al rendimiento deportivo**

| Relevancia frecuencia | Poco relevante | Muy relevante |
|---|---|---|
| **Poco frecuente** | **Poco frecuente y poco relevante**<br>Balón a tierra, bote neutral en fútbol. | **Poco frecuente y muy relevante**<br>Remates en fútbol. |
| **Muy frecuente** | **Muy frecuente y poco relevante**<br>Pases entre centrales sin presión en fútbol. | **Muy frecuente y muy relevante**<br>Pases en último 1/3 de campo en fútbol |

Fuente: Elaboración propia.

Se trata de sistematizar lo relevante.

En el ámbito general de la empresa, Charan (2007) identifica ocho criterios del "saber cómo" que pueden adaptarse al deporte:

- Identificar la(s) clave(s) más relevante(s) de la competición deportiva.

- Detectar amenaza(s) de la competencia.
- Liderar socialmente a los deportistas.
- Evaluar a jugadores y cuerpo técnico.
- Lograr la cohesión mediante acciones de *team-building*.
- Establecer objetivos.
- Establecer prioridades.
- Anticiparse al entorno y presión social.

## 2.4.3 Ventaja competitiva

Dos informes que consten de las mismas conclusiones y que muestren los mismos datos pueden diferenciarse en la forma de presentación y visualización que, al margen de los aspectos estéticos, logren llamar la atención del cuerpo técnico en uno de los casos mientras que en otro no lo haga. En esta perspectiva es interesante la diferenciación de visualizaciones que categoriza Jayal et al. (2018) en el deporte:

- Interactivas.
- Simulación.
- Modelado.
- Geo-espacial.
- Espacio-temporal.

¿Cuál será la próxima ventaja competitiva detectada que permitirá obtener victorias y éxitos? Existen casos concretos que ha dejado la historia del deporte, como el incremento del ritmo de juego en balonmano.

Podemos identificar casos de éxito en el transcurso de las décadas en el fútbol:

- **1920's:** profesionalismo.
- **1930's:** escuelas y sistemas de juego.
- **1950's:** talento en toma de decisiones.
- **1970's:** universalidad de puestos.
- **2000's:** transiciones.
- **2010's**: generación de superioridades numéricas o temporales.

En la mayoría de las ocasiones ha sido la intuición de jugadores o técnicos la que ha abierto el camino a innovaciones que marcaron una época de ventaja competitiva. En el otro lado de la balanza, se encuentran nuevas estrategias que partieron del análisis métrico de datos, como el aumento reciente de los intentos de lanzamientos triples, que tuvo como origen la concepción de Morey, ejecutivo desde 2007 de los Houston Rockets de NBA.

## Anexo IV

**Supuesto práctico de uso y transferencia de datos**

El desarrollador contratado utiliza el lenguaje de programación *Python* para transformar los archivos de *tracking* recibidos en el Campeonato Panamericano, y responder gráficamente las consultas realizadas por el cuerpo técnico sobre las distancias de las ayudas defensivas a medida que avanza el partido. Los archivos de *eventing* son tratados con el lenguaje R y soluciones SQL que permiten obtener las razones que decantan parciales temporales favorables o desfavorables. Sin embargo, no se disponen de los *feeds* (archivos originales de datos estandarizados) en las competiciones "domésticas" o torneos continentales de jóvenes, para lo que se realiza una aplicación Android que permite a los cuerpos técnicos el registro de un *eventing* básico de los propios partidos.

El proceso de creación de esta app sirve como excusa, en una de las visitas presenciales de Antonio, para conocer el *know-how* o saber hacer del director técnico y entrenadores de la Federación. Considera que el nivel de consultas sobre la fase defensiva del juego supera la que él ha conocido en Europa y debe incluso estudiar el juego con mayor profundidad para no quedarse atrás en el diseño de la app. Sin embargo, en las fases ofensivas (incluido el contraataque) el nivel de consultas es muy básico y estima, al consultar el *benchmarking* de los equipos del continente en las competiciones mundiales, que puede ser el pilar de una ventaja competitiva en el próximo ciclo olímpico.

Después de unos meses, Antonio regresa tras cruzar de nuevo el Atlántico y luego de la presentación de la app Android, aprovecha un momento de *quorum* para exponer los *rankings* de ataque, defensa, repliegue y contraataque de las selecciones en los últimos panamericanos y mundiales. La tendencia parece clara en los cuadros de mandos de Power BI, QlikView y Tableau que ha preparado: jugar rápido da ventaja en los últimos 5 torneos en la mayoría de partidos que ha jugado la selección. Los técnicos muestran matices pero sienten que es un *insight* que les convence para aumentar el rendimiento y la velocidad de juego de sus selecciones, que pueden verse favorecidas por el tipo de jugador del país.

Tanto las selecciones masculinas como femeninas aumentan su velocidad de juego en los siguientes torneos, mientras que en las selecciones masculinas se traduce en un aumento de pérdidas de balón y disminución del rendimiento, en la femenina se alcanzan niveles positivos nunca esperados. Se inicia una transición de jugadoras a ligas europeas que, en las concentraciones, refuerzan y matizan los detalles que permiten un aumento todavía mayor de la velocidad de juego con mejor precisión. Se consigue la clasificación para el mundial femenino mientras que las selecciones masculinas regresan a un ritmo más pausado de juego y mejoran levemente su rendimiento.

En una de las reuniones de presentaciones de memoria de la temporada, se presentan conclusiones sobre los motivos por los cuáles el aumento de la velocidad de juego ha sido “rentable” en los equipos femeninos, pero no en los masculinos. Se plantean hipótesis relacionadas con la escasa presencia de jugadores masculinos en las ligas europeas, mientras que ya hay más jugadoras femeninas en las ligas europeas que en la doméstica, y a su vez se encuentran ligeros matices en el criterio arbitral en las competiciones de mujeres y hombres.

Tras tomar nota de las hipótesis lanzadas por los técnicos, Antonio, mediante el programa SPSS, realiza un análisis de regresión de las variables disponibles en campeonatos Panamericanos. Pero son pocos partidos para obtener conclusiones, por lo que calcula el tamaño de la muestra necesario para realizar un informe presentable a la Federación.

Al realizar el cálculo, Antonio concluye que el número de partidos competitivos contra rivales que permitan mejorar a los equipos propios es muy escaso en los ciclos olímpicos. Recomienda aumentar el número de estos partidos, para lo cual es necesaria una mayor inversión, especialmente en equipos masculinos, cuya mayoría de jugadores no disputa ligas europeas.

No se dispone de presupuesto para llevar a cabo tal objetivo, pero la Federación decide establecer convenios con equipos de segundas divisiones europeas que permitan una mayor presencia de jugadores en estas ligas. Antonio realiza una estimación de la mejora de rendimiento que puede suponer la presencia de al menos 5 jugadores en estas competiciones.

Tras ejecutar el convenio, empiezan a llegar jugadores a las ligas europeas, que se agrupan en un principio en dos países limítrofes. Antonio es el responsable de realizar el seguimiento individual de estos jugadores, y es consciente que la mayoría de su carrera ha dedicado sus esfuerzos a datos colectivos y de equipo, pero no individuales. Debe formarse y exprimir el *know-how* (saber cómo) de los técnicos de la Federación en cuanto a aspectos individuales del juego, y diseñar nuevos instrumentos de registro. También realiza una "calculadora de mercado" que estima el precio que alcanza un jugador tras su estancia en Europa, tanto para un potencial traspaso del que la Federación obtiene derechos de formación, como del potencial sueldo que los clubes estarían dispuestos a proporcionar a los jugadores tras dos temporadas en segunda división.

Antonio debe calibrar y comunicar el margen de error a la Federación: existen casos de falsos positivos, es decir, jugadores a los que se estimó un alto rendimiento en la selección tras dos años en ligas europeas y que no mejoraron su rendimiento, y el opuesto de falsos negativos: jugadores descartados para participar en selecciones juveniles y absolutas, que alcanzaron el mercado europeo sin el apoyo de la Federación tras haber sido evaluados con los instrumentos de evaluación individual. En una de las reuniones, Antonio comunica que el número de falsos negativos se mantiene estable, pero que está aumentando el de falsos positivos, para lo que debe recalibrar la aplicación.

CONTINUE

# Referencias

**Alamar, B.** (2013). Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers [English edition]. New York: Columbia University Press.

**Charan, R**. (2007). Know-How: The 8 Skills that Separate People who Perform From Those Who Don't. Cornerstone.

**Davenport, T.** (2014a). Analytics in Sports: The New Science of Winning. SAS. Recuperado de https://www.sas.com/en_au/whitepapers/iia-analytics-in-sports-106993.htmlDavenport, T. (2014b) Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business Review Press, [240]. USA ©2014.

**Ford, P.** (2015). What is Code?. Bloomberg Businewssweek. Bloomberg

**Hohman, A., Lames, M. y Letzelter, M.** (2005). Introducción a la ciencia del entrenamiento. Barcelona: Paidotribo.

[Imagen sin título sobre Benjamin Alamar]. (s/f). Recuperado de https://www.linkedin.com/in/benalamar?trk=public_profile_browsemap_mini-profile_title.

[Imagen sin título sobre Paul DePodesta]. (2017). Recuperado de https://commons.wikimedia.org/wiki/File:Paul_DePodesta_(36380255290).jpg.

**Jayal, A., McRobert, A., y Oatley, G.** (2018). Sports Analytics (Routledge Studies in Sports Performance Analysis). Londres: Routledge.

**Lebed, F.** (2017). Complex Sport Analytics (Routledge Research in Sport and Exercise Science). New York: Routledge.

**O'Donoghue, P.** (2014). Data Analysis in Sport (Routledge Studies in Sports Performance Analysis). Londres: Routledge.

**Whitehead, A**. (1967). The Aims of Education and Other Essays. New York: The Free Press.

CONTINUE