

# Apuntes

## Simulacro 02 — Estrategia y Gestión Empresarial Basada en Análisis de Datos

**Fecha:** 18/02/2026 **Duración:** 120 minutos **Instrucciones:** Responda a los dos ejercicios. Dispone de aproximadamente una cara por ejercicio. Se valorará la capacidad de síntesis, la relación entre conceptos y la correcta interpretación analítica. Cuide la ortografía.

---

### Ejercicio 1 (5 puntos)

La operadora de telecomunicaciones **TelcoMóvil** tiene 200.000 clientes y ha detectado que su tasa de churn (abandono) ha subido del 8% al 14% en el último año. La empresa almacena datos en tres sistemas separados: un **CRM** (datos de contacto, antigüedad, tipo de contrato), un sistema de **facturación** (importes, consumo de datos, incidencias de pago) y una plataforma de **atención al cliente** (nº de reclamaciones, motivos, tiempo de resolución).

El Director General dice: *“Ya sabemos cuál es el problema: los clientes se van por precio. Hay que hacer un descuento del 20% a todos los que lleven más de un año. Así los retenemos.”*

El equipo de datos, analizando los datos del último año, ha descubierto que los factores que más influyen en el abandono son: el **número de reclamaciones** y **no tener contrato de permanencia**. Sin embargo, el **gasto mensual** y la **edad del cliente** no parecen influir significativamente en la decisión de irse.

**Se pide:**

**a)** Los datos de TelcoMóvil están en tres sistemas separados. Explique qué proceso y qué infraestructura necesitaría la empresa para poder realizar análisis integrados sobre sus clientes. ¿Qué problemas de calidad de datos podría encontrarse y cómo los resolvería?

b) Identifique qué tipo de decisión está tomando el Director General y explique sus riesgos. Contraste este enfoque con lo que ha descubierto el equipo de datos. Si el equipo quisiera además obtener **reglas sencillas** para que el call center identifique clientes en riesgo sin necesidad de consultar modelos, ¿qué técnica analítica propondría y por qué?

c) Basándose en los hallazgos del equipo de datos, diseñe una campaña de retención. Incluya: a qué clientes dirigirla (y por qué no a todos), qué tipo de oferta propone, cómo fijaría el umbral de exigencia para minimizar la dilución, y proponga dos KPIs SMART para medir el éxito.

---

## Ejercicio 2 (5 puntos)

La cadena de gimnasios **FitLife** (30.000 socios, 15 centros) ha realizado un análisis de segmentación mediante **clustering K-Means** sobre los datos de su tarjeta de socio. Se han identificado **4 clústeres**:

Variable	Clúster 1 (n=12.000)	Clúster 2 (n=3.500)	Clúster 3 (n=9.000)	Clúster 4 (n=5.500)
Antigüedad (meses)	4	42	18	30
Visitas/mes	1.8	12.5	3.2	7.4
Gasto mensual extra (€)	5	85	12	35
Clases grupales/mes	0.3	8.1	0.8	4.2
Tasa de cancelación últimos 6 meses (%)	35	2	28	8

Para explicar los clústeres al equipo comercial, el equipo de datos ha entrenado un **árbol de decisión** usando el clúster como variable objetivo. Las primeras reglas del árbol son:

Regla 1: Si Visitas/mes > 9.5 Y Gasto extra > 50€ → Clúster 2 (prob. 91%)

Regla 2: Si Visitas/mes ≤ 9.5 Y Antigüedad ≤ 8 meses Y Tasa cancelación > 25% → Clúster 1 (prob. 87%)

Regla 3: Si Visitas/mes ≤ 9.5 Y Antigüedad > 8 meses Y Clases grupales > 3.0 → Clúster 4 (prob. 82%)

**Se pide:**

a) Interprete los 4 clústeres: asigne un nombre descriptivo a cada grupo, justifique el perfil basándose en los datos numéricos de la tabla, y proponga una estrategia comercial diferenciada para cada segmento dentro del ciclo de vida del cliente.

b) Explique qué técnica se ha utilizado con el árbol de decisión y por qué. Interprete las tres reglas: ¿qué significan en lenguaje de negocio? ¿Por qué es útil esta combinación de modelos (clustering + árbol) para una empresa?

c) El departamento de marketing de FitLife propone tres iniciativas: 1. Compartir los datos de frecuencia cardíaca y peso corporal de los socios con una **compañía de seguros** asociada, a cambio de descuentos en las pólizas. 2. Rastrear la **geolocalización** de los socios a través de la app del gimnasio para enviarles notificaciones push cuando estén cerca de un centro. 3. Enviar emails personalizados con **recomendaciones de dieta** basadas en el historial de peso de los socios, sin haberles informado de este uso.

Identifique qué principios del RGPD se vulneran en **cada una** de las tres iniciativas y explique qué debería hacer la empresa para actuar correctamente.

---

**Nota para el alumno:** Recuerda citar los datos numéricos de las tablas cuando justifiques tus respuestas. Si una pregunta tiene varias partes, responde a todas.

---

## Respuestas modelo

---

### Respuesta modelo 1a — Infraestructura y calidad de datos

TelcoMóvil necesita un proceso **ETL** (Extract, Transform, Load) para integrar los tres sistemas en un **Data Warehouse (DTW)**:

- **Extract:** Extraer datos del CRM, del sistema de facturación y de la plataforma de atención al cliente. Cada sistema es una fuente operacional independiente (como un ERP o un CRM).
- **Transform:** Fase crítica donde se limpian y unifican los datos. Problemas típicos:
  - **Duplicados:** Un mismo cliente puede estar registrado con variaciones de nombre en los tres sistemas (“J. García”, “Juan García López”, “Juan García”). Se necesita construir un **Golden Record**: una versión única y veraz de cada cliente.
  - **Nulos e inconsistencias:** Campos vacíos, formatos de teléfono distintos, fechas en formatos diferentes.
  - **Datos desactualizados:** Direcciones antiguas, contratos ya renovados que aparecen como vigentes en un sistema pero no en otro.
- **Load:** Cargar los datos limpios y unificados en el DTW, que es un repositorio diseñado para análisis (no para operativa diaria).

Sin este proceso, la empresa no puede cruzar la información de reclamaciones con la de facturación y contrato. Es imposible saber, por ejemplo, que el cliente que ha reclamado tres veces este trimestre es el mismo que no tiene permanencia y tiene alta probabilidad de churn. **Sin Golden Record, tratas al mismo cliente como tres clientes distintos** y cualquier análisis posterior está contaminado.

---

## Respuesta modelo 1b — HIPPO vs Data-Driven + técnica analítica

El Director General está tomando una decisión de tipo **HIPPO** (Highest Paid Person's Opinion): se basa en su intuición ("se van por precio") sin evidencia analítica. Los riesgos son:

1. **Hipótesis incorrecta:** Los datos muestran que el gasto mensual NO es un factor significativo de abandono. El problema no es el precio, son las reclamaciones y la falta de permanencia. Un descuento del 20% no resolverá la insatisfacción por mal servicio.
2. **Destrucción de margen sin retorno:** Aplicar un 20% a todos los clientes de más de un año es una promoción masiva sin segmentar. Se produce **dilución**: se regala descuento a clientes que no se iban a ir (los que tienen contrato de permanencia, los que no reclaman). Eso es destruir margen gratis.
3. **No medible:** Sin objetivo data-driven, no se puede saber si la campaña funcionó o fracasó.

El enfoque **data-driven** es superior porque primero identifica las causas reales (reclamaciones, ausencia de permanencia), permite segmentar clientes por riesgo, y diseñar acciones dirigidas solo a los que realmente están en peligro de irse.

Para obtener **reglas sencillas** que el call center pueda aplicar, la técnica adecuada es un **árbol de decisión**. Es un modelo supervisado que genera reglas jerárquicas legibles, del tipo: "*Si el cliente ha puesto más de 2 reclamaciones en el trimestre y NO tiene contrato de permanencia → riesgo alto de churn.*" Estas reglas son directamente accionables por un agente telefónico sin necesidad de consultar modelos estadísticos. Se puede combinar con la técnica de **"engañar al ordenador"**: si previamente se ha hecho un clustering para agrupar clientes por comportamiento, el árbol traduce esos grupos en reglas comprensibles para negocio.

---

## Respuesta modelo 1c — Campaña de retención

**A quién dirigirla:** Al segmento de clientes con alta probabilidad de churn — aquellos que acumulan reclamaciones y no tienen contrato de permanencia. NO a todos, porque: - Los clientes con permanencia ya están vinculados contractualmente (bajo riesgo). - Los clientes sin reclamaciones no muestran señales de abandono. - Dar descuento a estos grupos sería **dilución** pura: destruir margen en clientes que se quedan igualmente.

**Tipo de oferta:** En lugar de un descuento en precio (que los datos dicen que no es el problema), proponer una oferta ligada a **resolver la causa real del churn**: por ejemplo, acceso a un canal prioritario de atención al cliente durante 6 meses, o resolución garantizada de incidencias en 24h. Si se quiere añadir un incentivo económico, condicionarlo a renovar contrato de permanencia (así también reduces el segundo factor de riesgo).

**Umbral de exigencia:** La oferta solo se activa si el cliente renueva permanencia por 12 meses. Esto asegura que el incentivo cambia el comportamiento (de "sin contrato" a "con contrato"), no que se regale a quien ya iba a quedarse. El threshold debe estar **por encima del comportamiento habitual** del segmento objetivo.

**KPIs SMART:** 1. “Reducir la tasa de churn del segmento sin permanencia y con reclamaciones del 14% al 9% en los próximos 6 meses” (mide retención directa). 2. “Aumentar la tasa de renovación de contrato de permanencia del segmento objetivo del 25% al 40% durante Q2 2026” (mide cambio de comportamiento).

---

## Respuesta modelo 2a — Interpretación de clústeres

**Clúster 2 — “VIP / Fans”** (3.500 socios, 12% de la base): Máxima antigüedad (42 meses), máximas visitas (12,5/mes), máximo gasto extra (85€), máxima participación en clases (8,1/mes) y mínima cancelación (2%). Son socios consolidados que viven el gimnasio. Están en fase de **retención/desarrollo**. Estrategia: **up-selling** — ofrecerles servicios premium (entrenador personal, planes de nutrición avanzados), programa de embajadores, acceso anticipado a nuevos centros o clases. No necesitan descuentos: ya están comprometidos. El objetivo es maximizar su valor de vida y usarlos como prescriptores.

**Clúster 1 — “Nuevos en riesgo”** (12.000 socios, 40% de la base): Menor antigüedad (4 meses), mínimas visitas (1,8/mes), mínimo gasto extra (5€), apenas clases grupales (0,3/mes) y la mayor tasa de cancelación (35%). Son socios recién llegados que no han enganchado con el gimnasio. Están en fase de **captación** y muchos se pierden antes de consolidarse. Estrategia: **activación y cross-selling** — onboarding personalizado (clase de prueba gratuita, sesión con monitor), incentivos condicionados a asistencia mínima (ej: si vienes 8 veces este mes, el siguiente mes tienes acceso a clases grupales gratis). El objetivo es crear hábito antes de que cancelen.

**Clúster 3 — “Dormidos / En riesgo de abandono”** (9.000 socios, 30%): Antigüedad media (18 meses), baja frecuencia (3,2 visitas/mes), bajo gasto extra (12€), pocas clases (0,8/mes) y alta tasa de cancelación (28%). Llevan tiempo pero casi no usan el gimnasio. Son socios que pagan la cuota pero no vienen — y muchos están decidiendo irse. Están en fase de **retención urgente**. Estrategia: campaña de **reactivación** — contacto proactivo preguntando por qué no vienen, oferta de clases grupales de prueba (su participación es casi cero: 0,8/mes), encuesta de satisfacción. Si el problema es horario, ofrecer flexibilidad. Si es motivación, programa de acompañamiento.

**Clúster 4 — “Regulares / En desarrollo”** (5.500 socios, 18%): Buena antigüedad (30 meses), frecuencia moderada-alta (7,4 visitas/mes), gasto medio (35€), participación notable en clases (4,2/mes) y baja cancelación (8%). Son socios estables que usan el gimnasio de forma regular pero no son “fans”. Están en fase de **desarrollo**. Estrategia: **cross-selling y up-selling** — incentivar que prueben más servicios (ya hacen 4,2 clases/mes, podrían llegar al nivel del Clúster 2 con el empujón adecuado). Oferta: packs de clases especializadas, descuento en nutricionista si mantienen frecuencia. El objetivo es convertirlos en VIP.

---

## Respuesta modelo 2b — “Engañar al ordenador” + interpretación de reglas

La técnica utilizada es lo que se conoce como “**engañar al ordenador**”: se usa un modelo **supervisado** (árbol de decisión) tomando como variable objetivo el resultado de un modelo **no supervisado** (la etiqueta de clúster asignada por K-Means).

**¿Por qué se hace?** El clustering agrupa socios por similitud estadística, pero no explica POR QUÉ están en ese grupo. El algoritmo K-Means no genera reglas legibles. El árbol de decisión traduce esos grupos en **reglas de negocio comprensibles**:

- **Regla 1:** “*Si un socio viene más de 9,5 veces al mes y gasta más de 50€ extra → es un VIP (91% de probabilidad).*” En lenguaje de negocio: los socios que más vienen y más gastan son los fans. La variable más discriminante es la frecuencia de visitas combinada con el gasto.
- **Regla 2:** “*Si viene poco ( $\leq 9,5$  visitas), lleva menos de 8 meses y tiene tasa de cancelación alta ( $> 25\%$ ) → es un Nuevo en riesgo (87%).*” En negocio: los recién llegados que no vienen y ya muestran señales de abandono. Son los que necesitan intervención inmediata.
- **Regla 3:** “*Si viene poco, pero lleva más de 8 meses y participa en más de 3 clases grupales al mes → es un Regular (82%).*” En negocio: los socios estables cuyo vínculo con el gimnasio son las clases grupales, no el uso libre de máquinas.

**Utilidad para la empresa:** El equipo comercial puede clasificar socios con estas reglas sin ejecutar ningún modelo. Un responsable de centro puede mirar la ficha de un socio (visitas, gasto, antigüedad) y saber inmediatamente en qué segmento cae y qué acción tomar. Además, permite detectar socios que están **cerca de cambiar de segmento** — por ejemplo, un Regular con 9 visitas/mes y 48€ de gasto está a punto de ser VIP: una pequeña acción podría acelerar esa transición.

---

## Respuesta modelo 2c — RGPD

### Iniciativa 1 — Compartir datos de salud con aseguradora:

Principios vulnerados: - **Limitación de finalidad:** Los datos de frecuencia cardíaca y peso se recogieron para el servicio del gimnasio (seguimiento de entrenamiento), no para evaluación de riesgo asegurador. Compartirlos con una compañía de seguros es un cambio de finalidad no autorizado — el mismo tipo de violación que el caso Cambridge Analytica (datos recogidos para un fin, usados para otro). -

**Consentimiento expreso:** Los datos de salud son **datos sensibles** (categoría especial según el RGPD). Requieren consentimiento explícito y reforzado para cualquier tratamiento, y específicamente para compartirlos con terceros. -

**Minimización:** La aseguradora no necesita el historial completo de frecuencia cardíaca para ofrecer un descuento. Se estarían compartiendo más datos de los necesarios.

Qué hacer: Informar al socio del nuevo uso, obtener consentimiento explícito específico para compartir con la aseguradora, limitar los datos compartidos al mínimo necesario (ej: solo un indicador de “activo/inactivo”, no las métricas de salud brutas), y permitir retirar el consentimiento en cualquier momento.

### **Iniciativa 2 — Geolocalización para notificaciones push:**

Principios vulnerados: - **Consentimiento expreso:** El rastreo de geolocalización requiere consentimiento explícito previo. No basta con que el socio tenga la app instalada. - **Minimización:** Rastrear la ubicación continuamente para enviar una notificación cuando pasa cerca es desproporcionado. Se recogen más datos de los necesarios para el fin comercial. - **Transparencia:** El socio debe ser informado de forma clara de que su ubicación está siendo monitorizada y con qué finalidad.

Qué hacer: Pedir permiso explícito en la app con explicación clara (“¿Quieres que te avisemos cuando estés cerca de un gimnasio FitLife?”), permitir desactivar la función en cualquier momento, y no rastrear cuando la función esté desactivada.

### **Iniciativa 3 — Recomendaciones de dieta sin informar:**

Principios vulnerados: - **Limitación de finalidad:** El historial de peso se recogió para seguimiento de entrenamiento, no para recomendaciones de dieta. Es un cambio de finalidad. - **Transparencia:** No se ha informado a los socios de que sus datos de peso se usarán para este fin. El RGPD exige que el interesado sepa qué tratamiento se hace con sus datos. - **Consentimiento expreso:** Los datos de peso son datos de salud (categoría especial). Su uso para un fin nuevo requiere consentimiento reforzado.

Qué hacer: Informar previamente a los socios del nuevo uso, obtener consentimiento opt-in explícito antes de enviar cualquier email, ofrecer la opción de exclusión, y asegurarse de que las recomendaciones las supervise un profesional (no solo algoritmos) dado que implican salud.