

Apuntes

Simulacro 01 — 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 empresa **FreshMarket**, una cadena de supermercados con 45 tiendas, quiere lanzar una campaña de fidelización. El Director Comercial propone, basándose en su experiencia de 20 años, invertir 200.000€ en un descuento del 15% en productos frescos para todos los clientes durante 3 meses, porque “los frescos es lo que más se vende y así vendrá más gente”.

El departamento de datos, recién creado, sugiere que antes de lanzar la campaña se analicen los datos de los últimos 12 meses del programa de tarjeta de fidelización.

Se pide:

a) Identifique qué tipo de toma de decisión representa la propuesta del Director Comercial. Explique los riesgos de este enfoque y argumente por qué el proceso analítico propuesto por el departamento de datos sería más adecuado.

Respuesta del alumno:

Esto corresponde con una forma de tomar decisión concodida como HIPPO (Hights paid person opinion). Supone un grave riesgo para la toma de decisiones porque al no seguir una metodología data-driven, las tomas de decisión no plantean investigación previa descriptiva, segmentación ni validación de escenarios hipotéticos. Esto supone un mayor riesgo de equivocarse y no tener un ROI positivo

Corrección 1a (Nota: 7/10):

Bien: Identificas HIPPO correctamente. Mencionas que falta investigación descriptiva, segmentación y validación de escenarios. Conectas con ROI negativo.

Mal: “Hights” → “Highest” (ortografía cuenta). Falta argumentar explícitamente por qué el enfoque data-driven es MÁS ADECUADO — la pregunta lo pide y solo dices que HIPPO es malo, sin desarrollar la alternativa.

Respuesta modelo 1a:

La propuesta del Director Comercial es un ejemplo de decisión **HIPPO** (Highest Paid Person's Opinion): se basa en la intuición y la experiencia subjetiva, no en evidencia analítica. Los riesgos son: (1) invertir 200.000€ sin validar si la hipótesis es correcta, (2) no poder medir por qué funcionó o fracasó, (3) aplicar una promo indiscriminada que destruye margen en clientes que habrían comprado igual (dilución).

El enfoque analítico es más adecuado porque permite **validar la hipótesis antes de invertir**: analizar los datos del programa de fidelización para entender qué clientes compran frescos, segmentar perfiles, simular escenarios de rentabilidad (BEP) y diseñar una acción dirigida. Además, permite medir resultados y corregir en tiempo real.

b) Describa, paso a paso, el proceso que seguiría el departamento de datos para fundamentar (o refutar) la propuesta del Director Comercial. Incluya: qué datos necesitaría, qué tipo de análisis realizaría y qué KPIs definiría para medir el éxito de la campaña (proponga al menos dos KPIs SMART).

Respuesta del alumno:

El equipo de datos necesitaría los datos de dimensiones de clientes y sucesos extraídos del data-warehouse. Una vez opténidos haría lo siguiente: 1- Análisis descriptivo: esto te da información de cómo se comportan los clientes de forma general. Utilizando medidas estadísticas podremos observar donde se encuentran las desviaciones y las medias en función de los datos demográficos y psicográficos de los clientes. Esto nos permite abordar diferentes hipotésis. En este caso queremos ver cómo se comportan aquellas personas específicamente los clientes que compran frescos y los que no compran frescos

2-Segmentación: Tendríamos que validar si aquellas personas que compran frescos siguen un patrón o varios patrones comunes. Por ejemplo, si tienen determinada edad, compran a determinada hora o además combinan su compra con otros artículos. Esta parte es crucial para tratar de dirigir la campaña a los clientes adecuados. Una promoción en si mismo es una destrucción de margen con la esperanza de cambiar el comportamiento de los consumidores a los que van dirigidos para aumentar su respuesta en ticket medio o en cantidad de clientes. Por eso esta parte es crucial, si los clientes a los que les entrego la oferta ya me iban a comprar frescos igualmente con la misma respuesta, estoy sencillamente destruyendo mi margen.

3- Análisis prescriptivo: Para validar la propuesta del director comercial haríamos una matriz de calculo BEP (Break Even Point). Es decir, tendríamos que simular los diferentes comportamientos plausibles utilizando las variables de Respuesta: Gasto (ticket medio) y Número de clientes. Si el margen generado por el alcance de las variables de respuesta en la matriz BEP de cálculo es plausible (pongamos que el alcance es un 1% para ambas variables), o si el margen generado es netamente alto aún siendo más exigentes con las variables de respuesta, entonces tendría sentido lanzar la oferta, dado que estamos minimizando riesgos de fracaso.

Para esta campaña podríamos proponer dos KPIS: “Aumentar la tasa de conversión en un 2% para el segmento de personas senior de productos frescos durante 2 meses frente al mismo periodo del año anterior”

“Aumentar la tasa de respuesta de base de clientes en un 2% para el segmento de personas junior para productos frescos durante 2 meses frente al mismo periodo del año anterior”

Corrección 1b (Nota: 7,5/10):

Bien: Mejora enorme respecto al primer intento. Estructura clara en 3 pasos (descriptivo, segmentación, prescriptivo). Menciona destrucción de margen y dilución. El BEP está bien explicado: matriz, variables r y g, plausibilidad. Los KPIS tienen formato SMART (segmento, métrica, %, periodo, referencia temporal).

Mal: Los dos KPIS son casi iguales: “tasa de conversión” y “tasa de respuesta” son prácticamente lo mismo en este contexto. Mejor uno de respuesta y otro de gasto/ticket medio. No mencionas qué datos concretos necesitas (transacciones, productos, fechas, datos sociodemográficos de la tarjeta). Ortografía: “opténidos” → obtenidos, “estádísticas” → estadísticas, “frecos” → frescos, “dirijidos” → dirigidos, “prescriptivo” → prescriptivo, “alcanze” → alcance.

Respuesta modelo 1b:

Datos necesarios: Del programa de tarjeta de fidelización extraeríamos del Data Warehouse: transacciones (fecha, importe, productos), perfil del cliente (edad, frecuencia, antigüedad) y categorías compradas. Estos datos deben pasar por un proceso ETL para garantizar calidad (Golden Record, sin duplicados).

Paso 1 — Análisis descriptivo: Entender el comportamiento actual: qué clientes compran frescos, con qué frecuencia, qué ticket medio, qué combinan con los frescos. Esto permite formular hipótesis sobre el impacto real de un descuento.

Paso 2 — Segmentación: Agrupar clientes por comportamiento de compra (clustering). No todos reaccionarán igual a la promo. Un cliente que ya compra 80€/mes en frescos no necesita un descuento — dárselo sería dilución pura.

Paso 3 — Simulación BEP: Para cada segmento, construir una matriz de escenarios cruzando incremento de respuesta (r) × incremento de gasto (g). Calcular en qué punto el margen incremental cubre el coste de la promoción. Si el BEP requiere incrementos irrealistas, la promo no es viable.

KPIs SMART: 1. “Incrementar el ticket medio en frescos del segmento ocasional de 18€ a 23€ durante los 3 meses de campaña” (mide gasto) 2. “Aumentar la frecuencia de compra de frescos en el segmento joven de 1,2 a 1,8 visitas/mes durante Q2 2026” (mide respuesta)

c) El departamento de datos descubre que hay 3 segmentos de clientes claramente diferenciados. Explique por qué una promoción idéntica para todos los clientes puede ser ineficiente y qué enfoque alternativo propondría.

Respuesta del alumno:

Cada uno de los segmentos se comportaría de forma diferente. Es importante tener en cuenta la dilución de la promoción; la campaña tiene que dirigirse a cambiar el comportamiento de los clientes. Por tanto el nivel de exigencia tiene que ser distinto para diferentes clientes para que puedan acceder a su “premio” Por ejemplo, para clientes senior que tienen un ticket medio de 50 € le aplicaríamos la promoción a partir de 60 €. Y para clientes junior a partir de 40 € en el caso que su ticket medio fuera de 20 €.

Corrección 1c (Nota: 7/10):

Bien: Menciona dilución. Propones exigencias distintas por segmento con ejemplos concretos y números. Entiendes que el threshold debe estar por encima del comportamiento habitual.

Mal: Falta explicar explícitamente POR QUÉ una promo igual es ineficiente: el BEP es distinto para cada segmento porque tienen distinto margen, ticket y frecuencia. Podrías mencionar que incluso el tipo de oferta puede ser distinto por segmento (no solo la exigencia). Ortografía: “dirijirse” → dirigirse, “tiket” → ticket.

Respuesta modelo 1c:

Una promo idéntica es ineficiente porque cada segmento tiene un ticket medio, una frecuencia y un margen distintos, lo que significa que el **BEP es diferente para cada grupo**. Un descuento del 15% puede ser rentable para un segmento con margen alto y BEP bajo, pero destruir valor en otro con margen ajustado. Además, se produce **dilución**: clientes que ya compran frescos habitualmente reciben un descuento innecesario.

Enfoque alternativo: (1) Diferente exigencia por segmento: si el cliente habitual gasta 50€, la condición para acceder a la promo sería 60€. Si el cliente nuevo gasta 20€, la condición sería 30€. (2) Diferente tipo de oferta: al cliente fiel se le puede ofrecer acceso anticipado a productos premium (up-selling), mientras que al nuevo se le incentiva con cross-selling para que explore categorías. (3) Calcular el BEP para cada binomio oferta-segmento con la matriz de escenarios antes de lanzar.

Ejercicio 2 (5 puntos)

Una empresa de e-commerce ha realizado un análisis de segmentación (clustering K-Means) sobre su base de 50.000 clientes. El resultado ha generado **3 clústeres** con las siguientes medias:

Variable	Clúster 1 (n=22.000)	Clúster 2 (n=8.000)	Clúster 3 (n=20.000)
Antigüedad (meses)	6	38	24
Gasto mensual (€)	15	120	45
Frecuencia compra/mes	1.2	5.8	2.1
Tasa devolución (%)	18	3	22
Nº categorías compra	1.5	6.2	2.8

Se pide:

a) Interprete los 3 clústeres: asigne un nombre descriptivo a cada grupo y justifique el perfil de cada uno basándose en los datos de la tabla. ¿Qué estrategia comercial diferenciada propondría para cada segmento dentro del ciclo de vida del cliente?

Respuesta del alumno:

Cluster 2 - Clientes Fieles: Estos son los clientes que más tiempo llevan comprando nuestros productos, que compran más horizontalmente de nuestro catálogo, aman nuestros productos dado que nos los devuelven y gastan más. Son fieles a nuestra marca. Estos clientes ya tienen el comportamiento ideal, se podría reforzar con estrategias de upselling; no solo queremos que nos compren mucho en volumen, también en margen. Cuanto mejor podamos optimizar el mix de productos con más margen más rentables serán. También pueden ser usados como beta-testers: son los mejores clientes para darte feedback sobre tus nuevas apuestas de marca.

Cluster 1 - Newies: Tienen el rendimiento más bajo en todas las variables. Son clientes nuevos, que compran pocos productos, pocas veces y los devuelven con más asuidad. Para este cluster necesitamos aumentar su vida, para ello haría una promoción enfocada al cross-selling: queremos que prueben más productos de diferentes categorías, de forma que puedan encontrar aquello que les interesa de nuestra marca, reduciendo la tasa de devolución y aumentando su antigüedad

Cluster 3 - Growth - Están entre el cluster 1 y 2. Este cluster es peligroso porque tiene una alta tasa de devolución, es muy posible que muchos de ellos esten a punto de marcharse, hay que fidelizarlos. Haría una campaña de cross-selling para tratar de cambiarles el mix.

Corrección 2a (Nota: 7/10):

Bien: Los perfiles están bien identificados: Fieles, Nuevos, Intermedios. Las estrategias son coherentes (up-selling, cross-selling, fidelización). La idea de “beta-testers” para Cluster 2 es creativa y válida.

Mal: Error importante: “dado que **nos** los devuelven” → debería ser “dado que **NO** los devuelven” (tasa devolución 3%, la más baja). Falta citar los NÚMEROS de la tabla para justificar cada perfil — el profesor quiere ver que lees los datos concretos. Cluster 3: dices “alta tasa de devolución” pero no mencionas el 22% concreto. La estrategia es vaga: “cambiarles el mix” sin explicar cómo. Ortografía: “asuidad” → asiduidad, “estrategias” → estrategias, “antigüedad” → antigüedad, “fidelizalos” → fidelizarlos.

Respuesta modelo 2a:

Clúster 2 — “VIP / Fieles” (8.000 clientes): Máxima antigüedad (38 meses), máximo gasto (120€/mes), máxima frecuencia (5,8 compras/mes), mínima devolución (3%) y compra en 6,2 categorías. Son clientes consolidados en fase de **desarrollo/retención**. Estrategia: up-selling (productos premium, mayor margen), programa de fidelización exclusivo, usarlos como embajadores o testers de nuevos lanzamientos.

Clúster 1 — “Nuevos / Exploradores” (22.000 clientes): Menor antigüedad (6 meses), menor gasto (15€), baja frecuencia (1,2) y alta devolución (18%). Están en fase de **captación**. Estrategia: cross-selling para ampliar categorías (solo compran 1,5 de media), ofertas de bienvenida con exigencia baja para fomentar la segunda y tercera compra. El objetivo es reducir la devolución y aumentar la retención.

Clúster 3 — “En riesgo” (20.000 clientes): Antigüedad media (24 meses) pero la tasa de devolución más alta de los tres grupos (22%), gasto moderado (45€) y solo 2,8 categorías. Llevan tiempo pero no están satisfechos. Están en riesgo de **churn**. Estrategia: investigar causas de devolución, campaña de retención con incentivo condicionado a compra sin devolución, encuesta de satisfacción.

b) Para validar la segmentación, el equipo de datos ha utilizado un árbol de decisión posterior. La primera regla del árbol dice: “*Si Gasto mensual > 80€ y Antigüedad > 24 meses → Clúster 2 (probabilidad 94%)*”. Explique qué técnica se está aplicando aquí, por qué se usa un modelo supervisado para explicar un resultado no supervisado, y qué utilidad tiene esto para el negocio.

Respuesta del alumno:

El cluster es una caja negra. El árbol de decisión nos permite averiguar cuáles son las variables que más impacto tienen sobre la no supervisión. Es un modelo logit.

Corrección 2b (Nota: 3/10):

Bien: “El cluster es una caja negra” es correcto: el clustering agrupa pero no da reglas legibles.

Mal: “Es un modelo logit” es INCORRECTO. Un árbol de decisión NO es una regresión logística (logit). Son modelos completamente distintos: el árbol genera reglas tipo “si $X > 80$ y $Y > 24 \rightarrow$ grupo Z”, mientras que la regresión logística genera probabilidades con coeficientes. No explicas POR QUÉ se usa un supervisado para explicar un no supervisado. No explicas la utilidad para el negocio (la pregunta tiene 3 partes y solo respondes una parcialmente).

Respuesta modelo 2b:

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

Se hace porque el clustering agrupa clientes por similitud estadística pero **no explica las reglas** de por qué están en ese grupo. El árbol de decisión traduce esos grupos en **reglas de negocio legibles**: “Si $\text{gasta} > 80\text{€}$ y $\text{lleva} > 24$ meses \rightarrow es un VIP con 94% de probabilidad.”

Utilidad para el negocio: Permite que el equipo comercial, sin conocimientos técnicos, pueda identificar y clasificar clientes con reglas simples y accionables. También permite detectar a clientes que están “a punto” de pasar de un segmento a otro (por ejemplo, un cliente con gasto de 75€ y 23 meses está cerca de ser VIP) y actuar proactivamente para acelerar esa transición.

c) El departamento de marketing quiere utilizar los datos de navegación web, historial de compras y geolocalización de los clientes para enviarles ofertas personalizadas por SMS sin haber informado previamente a los clientes de este uso. Identifique qué principios del RGPD se estarían vulnerando y qué debería hacer la empresa para actuar correctamente.

Respuesta del alumno:

Estaría vulnerando el principio de consentimiento expreso.

Corrección 2c (Nota: 2/10):

Bien: Mencionas consentimiento expreso, que es uno de los principios vulnerados.

Mal: La pregunta pide PRINCIPIOS (plural) y QUÉ DEBERÍA HACER LA EMPRESA. Solo mencionas un principio de los 3-4 que se vulneran y no propones ninguna medida correctiva. En un examen donde se valora la relación de conceptos, una respuesta de una línea para una pregunta de desarrollo es insuficiente.

Respuesta modelo 2c:

Principios del RGPD vulnerados:

1. **Consentimiento expreso:** Los clientes no han aceptado que sus datos se usen para envío de SMS comerciales. El consentimiento debe ser explícito, informado y previo.

2. **Limitación de finalidad:** Los datos de navegación y compras se recogieron para operar el e-commerce, no para marketing directo por SMS. Usar geolocalización para SMS personalizado es un cambio de finalidad no autorizado (mismo error que Cambridge Analytica: datos recogidos con un fin, usados para otro).
3. **Minimización:** La geolocalización puede no ser necesaria para enviar ofertas. Se están usando más datos de los estrictamente necesarios para el fin comercial.
4. **Transparencia:** Los clientes no fueron informados de este tratamiento de sus datos.

Qué debería hacer la empresa: - Informar a los clientes del nuevo uso previsto (transparencia) - Obtener consentimiento explícito opt-in antes de enviar SMS (no basta con “si no rechaza, acepta”) - Evaluar si la geolocalización es realmente necesaria para la oferta (minimización) - Permitir al cliente retirar su consentimiento en cualquier momento - Documentar el tratamiento en el registro de actividades (Business Compliance)

Nota global orientativa

Apartado	Nota	Comentario
1a	7/10	Bien pero falta argumentar la alternativa data-driven
1b	7,5/10	Gran mejora. BEP bien integrado. KPIs casi correctos
1c	7/10	Buen concepto de dilución y exigencia diferenciada
2a	7/10	Buenos perfiles, falta citar números de la tabla
2b	3/10	Confundes árbol de decisión con regresión logística
2c	2/10	Muy corto. Un solo principio de 4, sin medidas correctivas
TOTAL	~5,5/10	Aprobado justo

Lagunas a trabajar antes del siguiente simulacro

1. **Árbol de decisión vs Regresión logística:** Son modelos distintos. Dominar el concepto de “engañar al ordenador” (supervisado para explicar no supervisado).
2. **RGPD:** Repasar los principios (limitación finalidad, minimización, transparencia, consentimiento) y practicar identificar VARIOS en un caso. Usar el diccionario en Resúmenes/diccionario_examen.md.
3. **Ortografía:** grave (no grabe), dirigidos (no dirijidos), obtenidos (no opténidos), asiduidad (no asuidad), prescriptivo (no prescritivo), antigüedad (no antiguedad).
4. **Citar datos:** Cuando interpretas una tabla, incluir los números concretos como evidencia.

5. **Respuestas completas:** Si la pregunta tiene 3 partes, responder las 3. Una línea no basta para una pregunta de desarrollo.