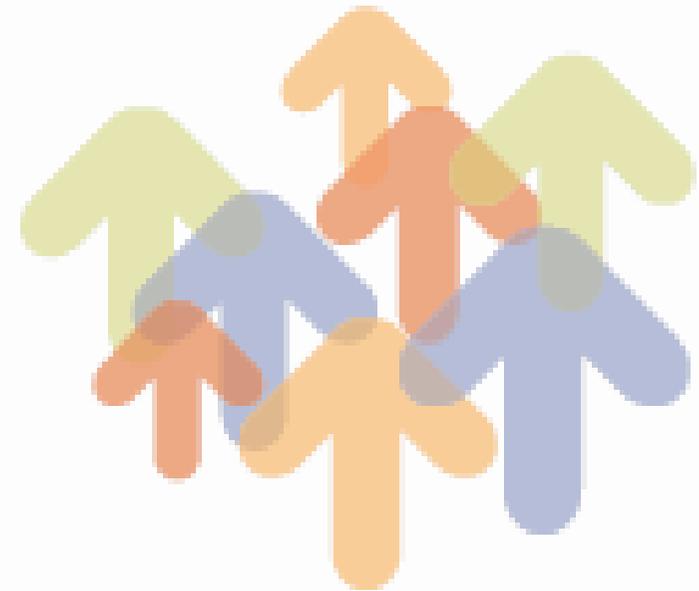


IBM Software Group

Aplicación de técnicas de minería de datos para la fidelización y retención de clientes

Víctor López Fandiño
Business Intelligence Solutions Specialist
IBM Software Group



ON DEMAND BUSINESS™

IBM Software Expo 2005

© 2005 IBM Corporation

Agenda

Puntos a tratar



- 1. El valor que aporta la minería de datos para el negocio**
 - ▶ Posicionamiento de la minería de datos: nuevas tendencias
 - ▶ Modelización de indicadores clave para el negocio
- 2. Cómo plantear un modelo retención de clientes de forma efectiva**
 - ▶ El problema de negocio y su impacto
 - ▶ Metodología de análisis y mejores prácticas
 - ▶ El cierre del ciclo analítico y sus beneficios
- 3. Visión integral para el despliegue de modelos**
 - ▶ Integración de la solución con los procesos de negocio
 - ▶ Infraestructura necesaria para la implementación
 - ▶ Fidelización mediante personalización de contenidos: un ejemplo práctico



Agenda

Puntos a tratar



- 1. El valor que aporta la minería de datos para el negocio**
 - ▶ **Posicionamiento de la minería de datos: nuevas tendencias**
 - ▶ **Modelización de indicadores clave para el negocio**
- 2. Cómo plantear un modelo retención de clientes de forma efectiva**
 - ▶ El problema de negocio y su impacto
 - ▶ Metodología de análisis y mejores prácticas
 - ▶ El cierre del ciclo analítico y sus beneficios
- 3. Visión integral para el despliegue de modelos**
 - ▶ Integración de la solución con los procesos de negocio
 - ▶ Infraestructura necesaria para la implementación
 - ▶ Fidelización mediante personalización de contenidos: un ejemplo práctico



De los datos a las decisiones a través de modelos

Los motores analíticos aportan una ventaja competitiva



- Cualquier iniciativa de modelización de los datos sobre clientes debe estar siempre encaminada a soportar la toma de decisiones estratégicas:



Datos



Conocimiento



Decisiones

patrones, tendencias, hechos, relaciones, asociaciones, predicciones

- Transacciones
- Sociodemográficos
- Respuesta a campañas
- Compra de productos
- Hábitos de comportamiento
- Quejas y reclamaciones
- Clasificaciones externas

- Segmentación de clientes
- Afinidad entre productos
- Patrones secuenciales de compra
- Rentabilidad de productos y clientes
- Rendimiento de campañas

- Tratamiento diferenciado a clientes
- Esquemas de precio diferenciados
- Optimización de la labor comercial
- Venta cruzada de productos
- Efectividad de promociones
- Incremento de cuota de mercado
- Retención de clientes
- Recuperación de clientes
- Captación de clientes

OLAP y minería de datos

Integración del análisis multidimensional con las técnicas de descubrimiento



- Dos enfoques para necesidades de negocio diferentes, pero que necesariamente deben complementarse

Análisis multidimensional

- Navegación sobre dimensiones y jerarquías del negocio
- Capacidad para modificar y reejecutar consultas con tiempos de respuesta inmediatos
- Capacidad para construir escenarios de simulación (*what if*)
- Implementación de modelos de cálculo complejos
- Capacidad para acceder a información de detalle, incluso en tiempo real



Minería de datos

- Capacidad para extraer tendencias, patrones, relaciones y desviaciones sobre grandes volúmenes de datos
- Resolución de problemas no abordables mediante técnicas de consultas
- Capacidad para aplicar modelos en tiempo real (entornos operacionales)
- Punto neurálgico dentro del CRM analítico

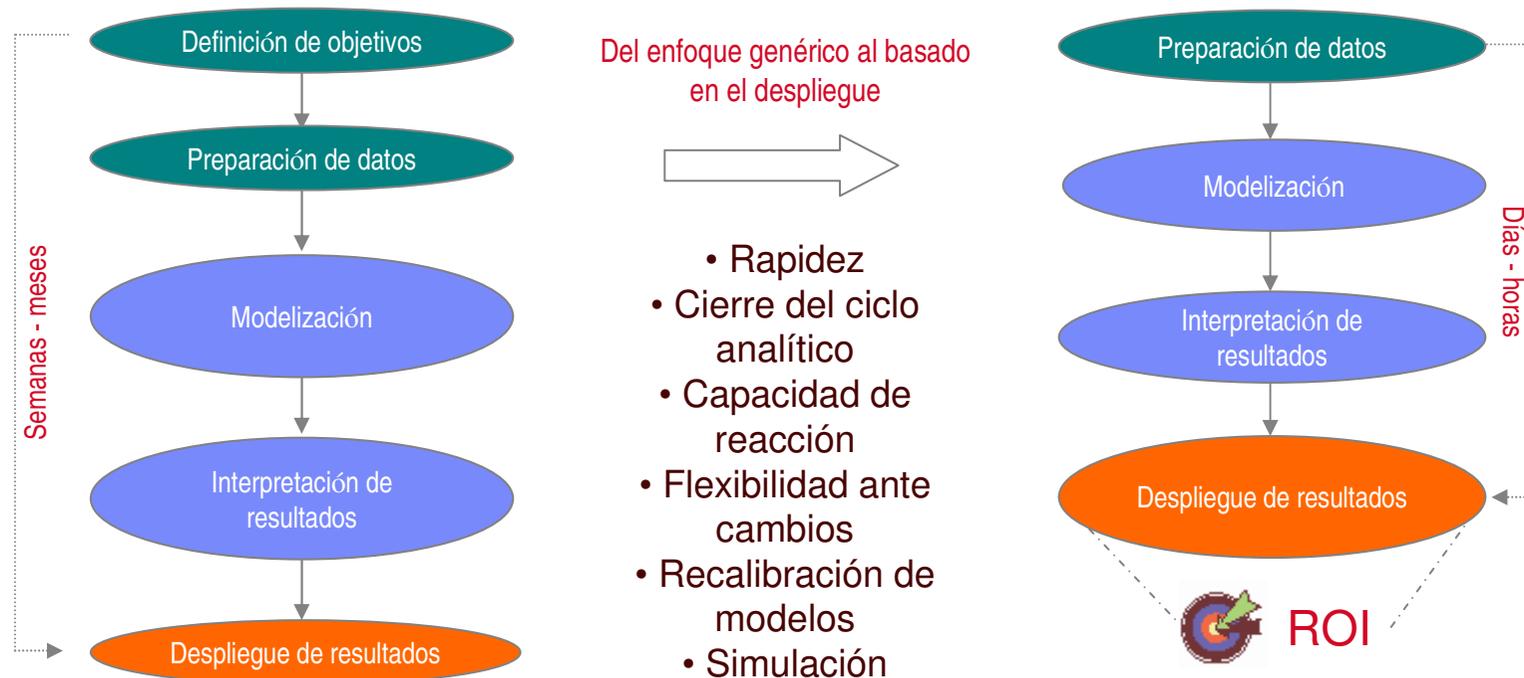


El reto actual: agilizar el tiempo de respuesta

Aplicar modelos flexibles en el punto de contacto con el cliente



- Desarrollo de modelos con suficiente flexibilidad y rapidez, aún a costa de perder precisión, no eficacia

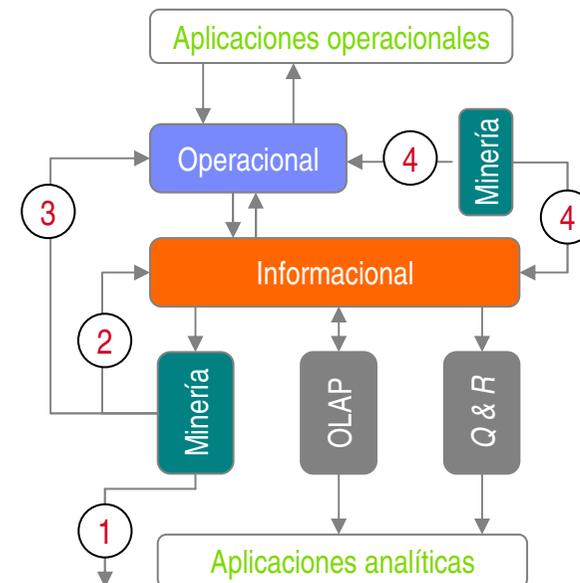


El cierre del ciclo analítico

¿Dónde aporta más valor de la minería de datos?



- Hay tres estadios relativos a la forma en que las empresas utilizan los análisis realizados mediante técnicas de minería de datos
 - 1. Uso aislado.** La minería de datos es un fin en si mismo (no hay integración)
 - 2. Integración en el entorno analítico.** Los resultados de los modelos se alimentan a otros sistemas, dentro del entorno informacional
 - 3. Cierre del ciclo analítico.** Los modelos son exportados a los entornos operacionales, donde se aplican según demanda y en tiempo real
 - 4. Automatización del ciclo analítico.** La generación, mantenimiento y exportación de los modelos a los entornos operaciones está automatizada

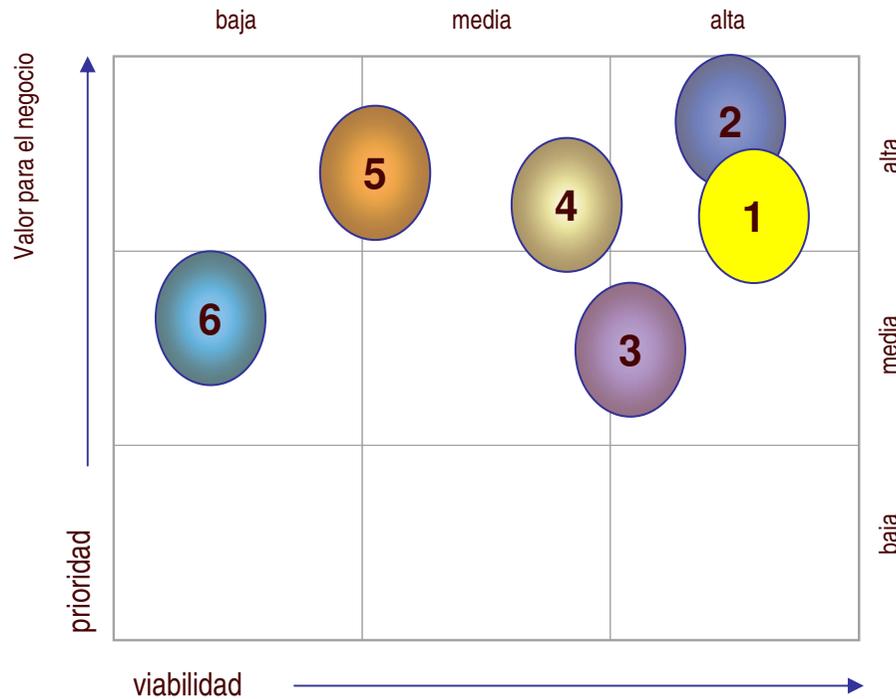


Modelización de indicadores dentro del negocio

Generación de métricas en función de la disponibilidad y viabilidad



- Es necesario planificar y abordar un esquema de análisis basado en una metodología coherente y secuencial



(Prioridades basadas en un caso práctico)

1. Segmentación de clientes
2. Asociaciones de productos
3. Perfilación de tiendas
4. Desactivación
5. Potencial del cliente
6. Respuesta a campañas

Pequeños consumidores	Clientes tradicionales
Grandes cuentas	Nuevos clientes



Los datos de partida necesarios

El éxito del análisis radica en la calidad de los datos fuente



- Para modelizar el comportamiento del cliente es necesario disponer de información histórica que permita medir su evolución dentro del negocio:

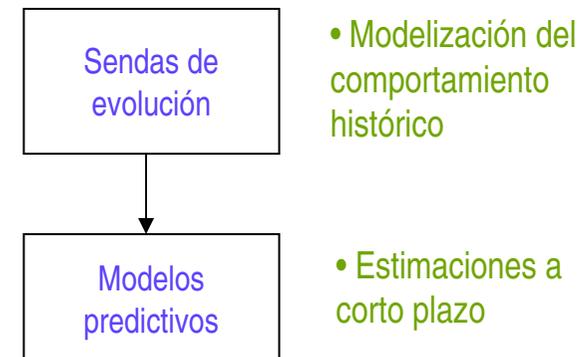


La información necesaria debe incluir:

- Datos sociodemográficos del cliente (identificables)
- Datos de comportamiento (frecuencia, importes, amplitud de compra...)
- Métricas de rentabilidad
- Estimaciones de su propensión de abandono y potencial a medio plazo

Para desarrollar una buena modelización, es necesario:

- Cierta profundidad de datos históricos como base para los modelos
- Información histórica sobre campañas, promociones y otras actividades de marketing.
- Histórico de los esquemas de precios y tarificación aplicados en cada caso
- Una buena calidad de datos



Impacto del grado de identificación de los clientes

La clave está en la trazabilidad



- El grado de conocimiento del cliente depende del nivel de detalle disponible sobre sus datos, lo que condiciona cualquier investigación:

Tipo de análisis	Cliente indentificados	Cientes rastreables	Cientes anónimos
• Segmentación de clientes (att. sociodemográficos)	√		
• Segmentación de clientes (hábitos de compra)	√	√	
• Segmentación de tipos de transacciones	√	√	√
• Segmentación de tipos de visitas	√	√	
• Segmentación de tiendas (perfiles de transacciones)	√	√	√
• Segmentación de tiendas (hábitos de compra)	√	√	
• Segmentación de tiendas (att. sociodemográficos)	√		
• Análisis de carros de la compra (transacciones)	√	√	√
• Asociación de compra a lo largo del tiempo	√	√	
• Relación de productos con rasgos sociodemográficos	√		
• Relación de productos con el tipo de pago	√	√	√
• Respuesta a campañas de márketing	√	√?	√?
• Impacto de promociones y descuentos	√	√?	

Agenda

Puntos a tratar



1. El valor que aporta la minería de datos para el negocio
 - ▶ Posicionamiento de la minería de datos: nuevas tendencias
 - ▶ Modelización de indicadores clave para el negocio
2. **Cómo plantear un modelo retención de clientes de forma efectiva**
 - ▶ **El problema de negocio y su impacto**
 - ▶ **Metodología de análisis y mejores prácticas**
 - ▶ **El cierre del ciclo analítico y sus beneficios**
3. Visión integral para el despliegue de modelos
 - ▶ Integración de la solución con los procesos de negocio
 - ▶ Infraestructura necesaria para la implementación
 - ▶ Fidelización mediante personalización de contenidos: un ejemplo práctico



Los principales retos analíticos en todo sector

El tipo de problemas que los paquetes integrados no pueden resolver



- Toda organización necesita tener un profundo conocimiento de cada cliente con el fin de comprender su dinámica y resolver los siguientes problemas:
 - Las iniciativas de perfilación de clientes son demasiado rígidas, con poco significado y complicadas de trasladar a los sistemas de producción y entornos transaccionales
 - El abandono de clientes se sitúa entorno al 15–20% anual. Existe una necesidad imperativa de reducir el abandono, desde el momento en que es más caro captar un nuevo cliente que retener a un existente (y rentable)
 - Entorno al 80% de los clientes de la mayoría de las organizaciones aportan una rentabilidad nula o negativa al negocio
 - El desarrollo de ofertas personalizadas para la adquisición y retención de clientes, a través de diseño de campañas basadas en patrones sociodemográficos y transaccionales, es complejo y costoso



Retención de clientes

Análisis del comportamiento del cliente con el fin de retenerlo y fidelizarlo



- Las iniciativas para retención de clientes deben estar basadas en un entorno analítico que permita un posterior despliegue de resultados en el negocio
 - El entorno analítico debe estar centrado en el estudio del abandono, analizando su impacto, causas y soluciones. El ROI de estas soluciones es elevado en un corto periodo de tiempo
 - Debe estar dirigido a analistas de negocio, principalmente de los departamentos de márketing, pero cubriendo necesidades operacionales en toda la organización
 - Debe contemplar un modelo de datos predefinido (pero flexible) orientado a la dinámica de clientes, con particularizaciones dependiendo del sector
 - Debe proporcionar capacidades de modelización integradas y de fácil utilización. La minería de datos es esencial para la perfilación y predicción del abandono, especialmente cuando se integra con motores OLAP
 - La experiencia del usuario dentro del entorno debe ser altamente personalizada y adaptable a cada perfil (tecnología J2EE basada en portal)



¿Quién analiza qué, donde y cuando?

Distintos roles en el negocio, distintas necesidades



- La retención y fidelización de clientes es una práctica que debe estar extendida por toda la organización, especialmente en el punto de contacto con el cliente
 - **Analista de minería de datos**
 - ▶ Definen, construyen y validan modelos de comportamiento
 - ▶ Generan contenidos para el resto de áreas de negocio
 - **Planificadores de campañas / ejecutivos de negocio**
 - ▶ Desarrollan estrategias de retención en función de los perfiles de los clientes
 - ▶ Monitorizan y estudian la evolución del abandono a lo largo del tiempo
 - **Responsables de CRM a nivel de cliente**
 - ▶ Reciben iniciativas de retención con el fin de personalizarlas y aplicarlas
 - ▶ Definen grupos localizados de clientes para acciones comerciales
 - ▶ Detectan, personalizan y aplican las estrategias de retención a sus clientes
 - **Personal en el punto de contacto**
 - ▶ Trabajan con vistas integradas de información del cliente, en tiempo real
 - ▶ Reaccionan y responden a los requerimientos del cliente, desplegando las iniciativas planteadas



Abandono de clientes

Construcción y despliegue de alarmas



- El análisis de abandonos permite construir un marco predictivo capaz de identificar la propensión de los clientes a cancelar determinados productos, o bien la totalidad de ellos.

- Hay que analizar el abandono en dos fases:

- Sendas de desvinculación

Rutas secuenciales de los clientes desde que entran en la compañía hasta que, potencialmente, abandonan.

- Predicción a medio plazo

Estimar la propensión de los clientes a cancelar determinados productos dentro de un periodo de tiempo determinado

- ... lo que permite:

- Comprender la actitud de los clientes.
- Identificar su propensión a abandonar.
- Actuar con el fin de reforzar su lealtad hacia la compañía antes de ser tentado por la competencia.

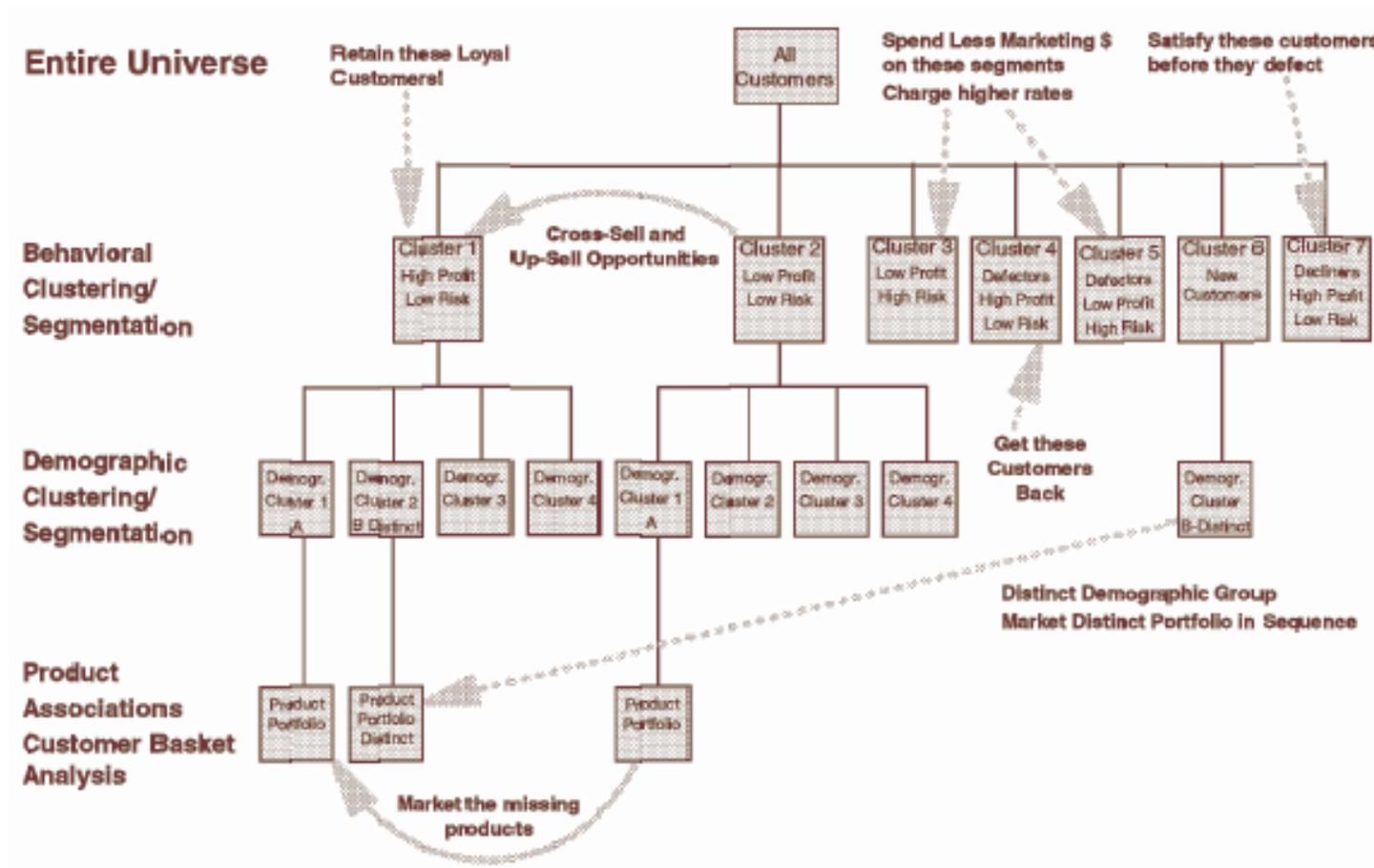


El punto de partida: segmentación de clientes

Perfilar al cliente de forma correcta



- La importancia e información que aporta un modelo de segmentación

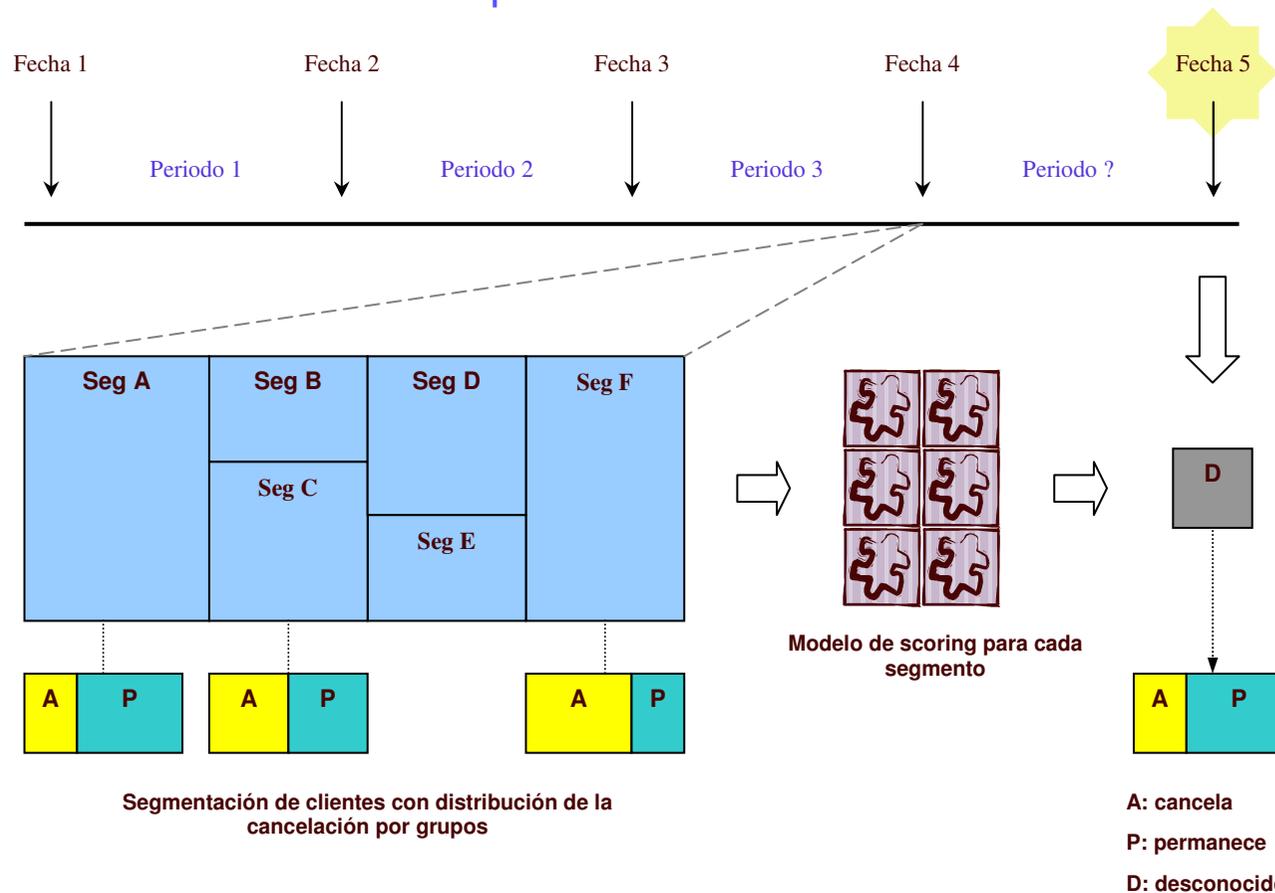


Esquema de segmentación temporal

Cómo evoluciona el perfil del cliente a lo largo del tiempo



- El análisis se basa en un esquema de segmentación a lo largo del tiempo, estudiando la evolución temporal de los clientes:

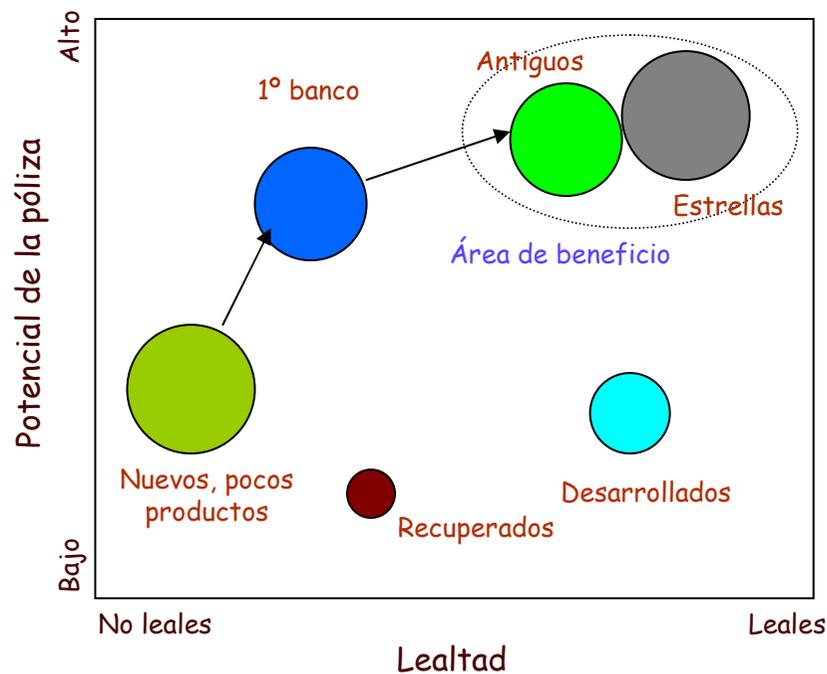


Evolución del cliente

Qué condiciones se dan para que el cliente abandone



- La segmentación de clientes no sólo permite identificar grupos homogéneos de estos, sino verificar y estudiar la evolución de su comportamiento en el tiempo:



El estudio de migraciones de clientes entre segmentos permite:

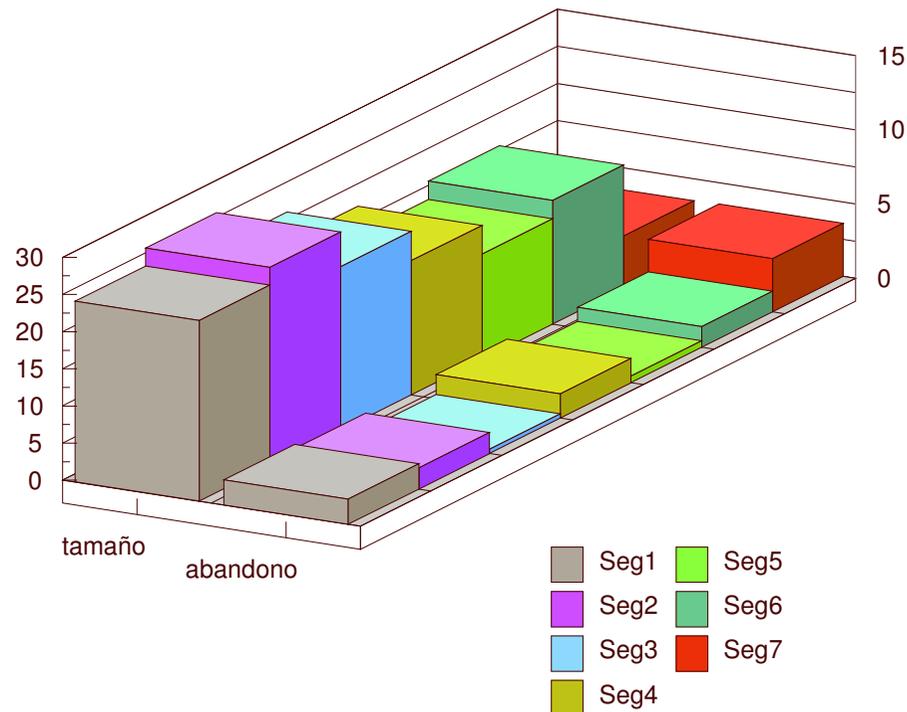
- Una visión completa de las dimensiones y movimientos de todos los segmentos
- Registrar la trazabilidad de los movimientos, permitiendo un análisis de tendencias a lo largo del tiempo
- Diseñar estrategias para guiar la migración hacia zonas de interés
- Enfocar los esfuerzos de retención, promociones y descarte de perfiles no interesantes

Distribución del abandono según perfiles

Centrarse en grupos rentables y con alta tasa de rotación



- El estudio de distribución del abandono por segmento permite correlacionar su ocurrencia con otros atributos de negocio y sociodemográficos



Descriptivas

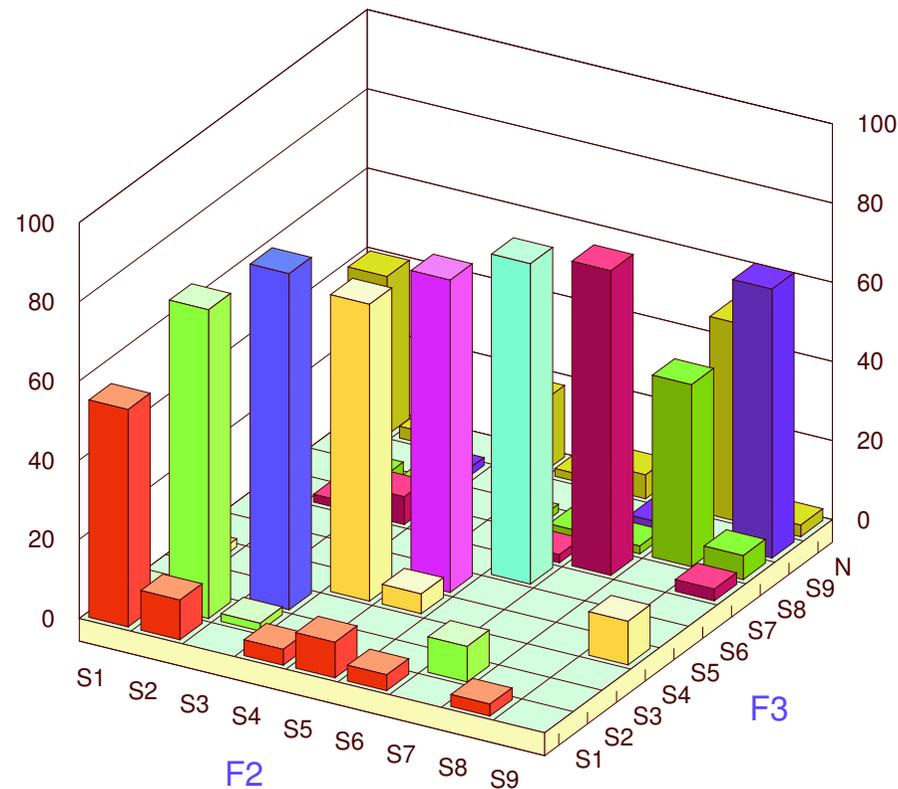
- Variables de discriminación respecto a la población global
- Distribución del abandono por segmentos
- Identificación de segmentos en términos de negocio
- Detección de segmentos con alto valor añadido (nichos)
- Correlación del abandono con indicadores

Sendas de desvinculación

Hacia donde van los clientes



- Un esquema de segmentación debe validarse a lo largo del tiempo; el comportamiento de los clientes debe ser consistente

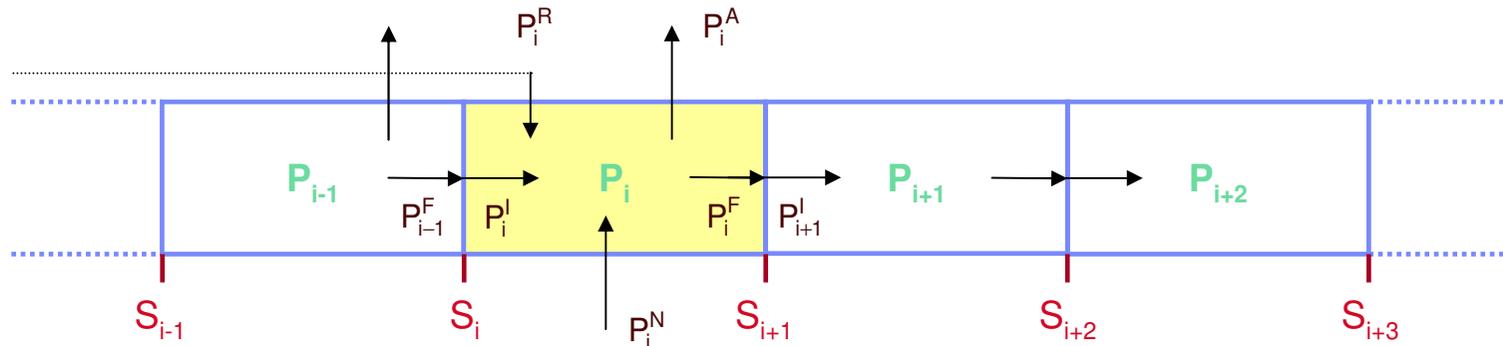


Análisis

- Estudio de la movilidad entre segmentos y validación de ésta en términos de negocio
- Determinación de patrones secuenciales de comportamiento y de posibles alarmas respecto al abandono
- Estudio del ciclo de vida y evolución de la nueva contratación
- Detección y calibración de las sendas de desvinculación según segmento

Dinámica de clientes

Cómo evolucionan los clientes a través de los periodos temporales



P_i^I : population at the beginning of P_i
 P_i^R : population reactivated during P_i
 P_i^N : new population during P_i
 P_i^A : population ceased during P_i
 P_i^F : population at the end of P_i
 S_i : population segmentation at P_i
 s_i^k : segment k in S_i ; $P_i^I = \sum_k s_i^k$

- Balance de clientes en un periodo

$$P_i^F = (P_i^I - P_i^A) + P_i^R + P_i^N = P_{i+1}^I$$

$$P_i^I + P_i^R + P_i^N = P_{i+1}^I + P_i^A$$

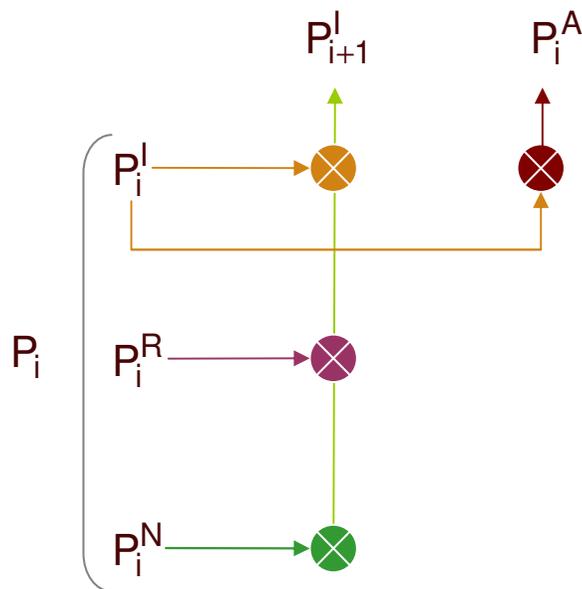
$$\left(\sum_k s_i^k \right) + (P_i^R + P_i^N) = \left(\sum_k s_{i+1}^k \right) + (P_i^A)$$

Análisis de rutas de evolución temporal

Hacia donde van los clientes, de donde vienen

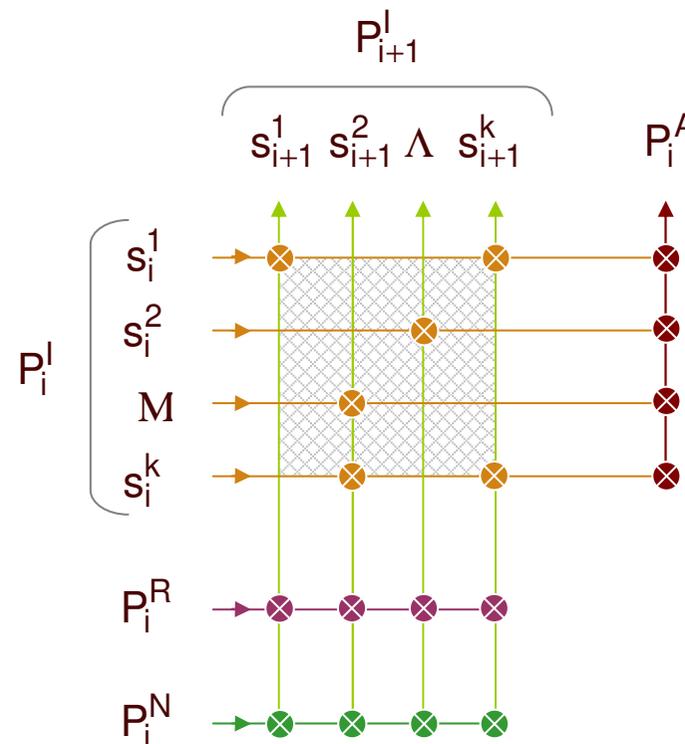


- Transiciones globales de clientes



Clientes nuevos y reactivados en un periodo, no pueden abandonar en el siguiente periodo (según la definición del indicador de abandono)

- Transiciones por segmento de cliente



Estimación del abandono a corto plazo

Quien va a abandonar y cuando

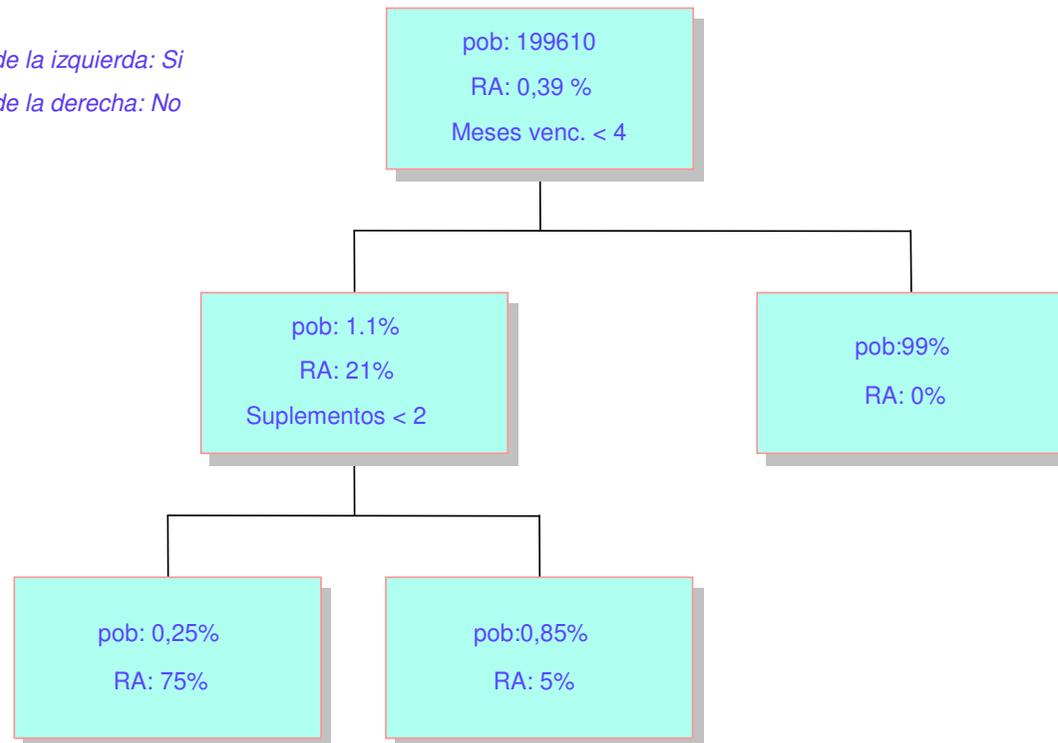


- Un modelo de clasificación en árbol permite obtener las reglas de asignación de un cliente a un grupo con una tasa de abandono determinado

S1

Rama de la izquierda: Si

Rama de la derecha: No



Reglas de asignación

- Potencia de predicción
- Posible solución

Potencial del cliente

Modelo de valor del cliente en el tiempo (LTV)



- El potencial indica la rentabilidad esperada para cada cliente. Permite identificar aquellos segmentos más rentables

$$\sum_k M(t+k)P(t+k)$$

- Valor actual del margen neto generado por cada cliente durante su vida económica

- M(t+k) es el margen esperado en el momento t+k
- P(t+k) es la probabilidad de permanencia como cliente en el momento t+k
- Este indicador permite ordenar la base de clientes en función del valor del margen esperado. Esto es útil para:
 - Identificación de los mejores clientes.
 - Optimización de las campañas y promociones de cara a retener a los mejores clientes.
 - Optimización de esfuerzos con clientes poco rentables.

Potencial del cliente

Modelo de valor del cliente en el tiempo (LTV)



- Como paso previo a la estimación del potencial, es necesario efectuar un análisis de la estabilidad temporal de la ordenación

- Ordenación estable:

Enero-Febrero			Marzo-Abril		
Pos.	Nome cliente	MARGINE	Pos.	Nome cliente	MARGINE
1	Mario Rossi	L. 1500	1	Mario Rossi	L. 1721
2	Giuseppe Verdi	L. 1359	2	Giuseppe Verdi	L. 1483
3	Antonio Bianchi	L. 1326	3	Antonio Bianchi	L. 1411
:			:		
:			:		
:			:		
98	Giorgio Ferrari	L. 286	98	Stefano Esposito	L. 312
99	Stefano Esposito	L. 252	99	Giorgio Ferrari	L. 291
100	Andrea Capelli	L. 174	100	Andrea Capelli	L. 198

- Si se excluyen previamente los abandonos, es probable obtener ordenaciones estables en el tiempo:

- En este caso, el empleo de técnicas predictivas convencionales (medias móviles y suavización exponencial) produce buenos resultados

- Ordenación inestable:

Enero-Febrero			Marzo-Abril		
Pos.	Nome cliente	MARGINE	Pos.	Nome cliente	MARGINE
1	Mario Rossi	L. 1500	1	Stefano Esposito	L. 1872
2	Giuseppe Verdi	L. 1359	2	Gennaro Caputo	L. 1513
3	Antonio Bianchi	L. 1326	3	Andrea Capelli	L. 1444
:			:		
:			:		
:			:		
98	Giorgio Ferrari	L. 286	98	Mario Rossi	L. 393
99	Stefano Esposito	L. 252	99	Giorgio Ferrari	L. 289
100	Andrea Capelli	L. 174	100	Antonio Bianchi	L. 198

- Si pese a excluir los abandonos la ordenación resulta inestable, hay que elaborar modelos más complejos:

- Segmentación de la población y cálculo de estimadores para cada segmento
- Predicciones neuronales considerando el modelo de abandono previo

Agenda

Puntos a tratar



- 1. El valor que aporta la minería de datos para el negocio**
 - ▶ Posicionamiento de la minería de datos: nuevas tendencias
 - ▶ Modelización de indicadores clave para el negocio
- 2. Cómo plantear un modelo retención de clientes de forma efectiva**
 - ▶ El problema de negocio y su impacto
 - ▶ Metodología de análisis y mejores prácticas
 - ▶ El cierre del ciclo analítico y sus beneficios
- 3. Visión integral para el despliegue de modelos**
 - ▶ Integración de la solución con los procesos de negocio
 - ▶ Infraestructura necesaria para la implementación
 - ▶ Fidelización mediante personalización de contenidos: un ejemplo práctico

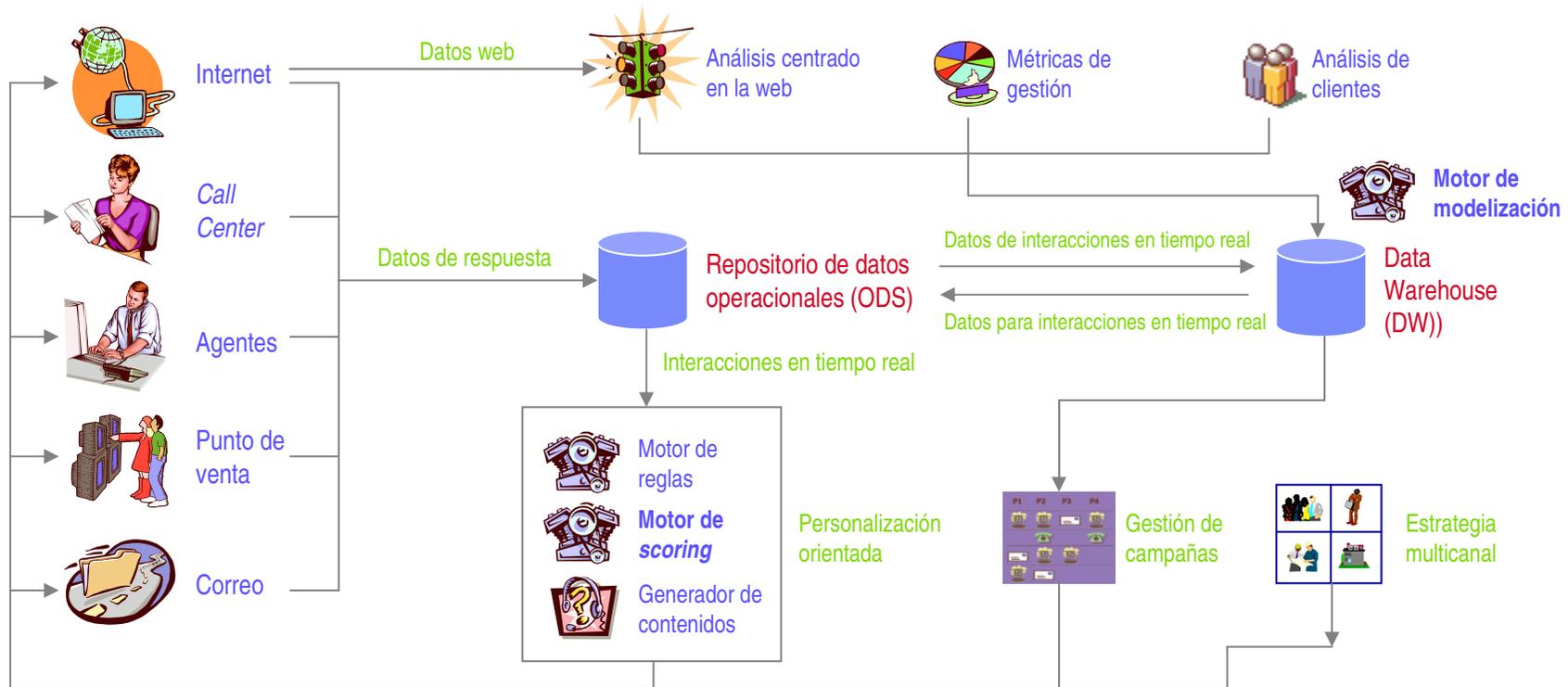


Integración de la información multicanal

La minería de datos debe implementarse como un conjunto de servicios



- Llevar los modelos del CRM analítico al operacional, de forma que todos los procesos de negocio puedan utilizarlos

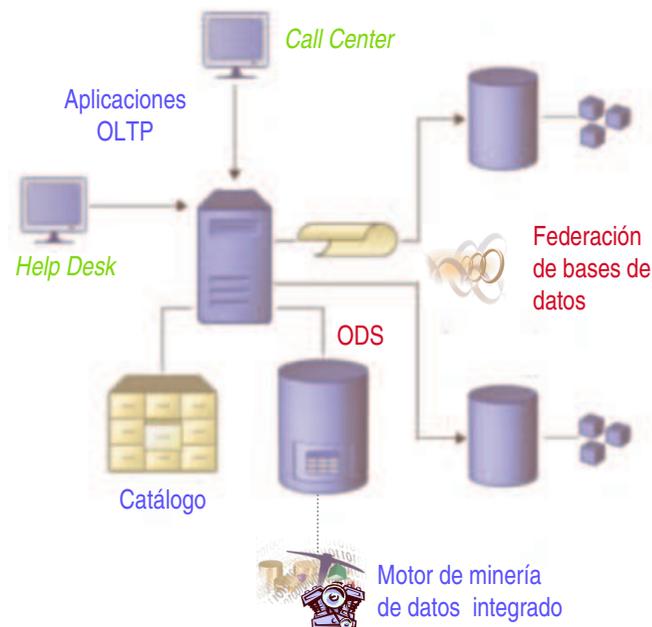


Componentes para una modelización efectiva

Rendimiento, transparencia, integración, automatización



- Funciones de minería de datos dentro del gestor de base de datos relacional (*in-database-data-mining*)



- **Facilidad de automatización e integración:** generación e invocación de modelos via SQL.
- **Eficacia operativa:** recalibración y actualización de modelos transparente.
- **Alto rendimiento:** el gestor construye y aplica los modelos, sin necesidad de mover datos.
- **Federación de fuentes de datos:** un modelo se aplica a datos dispersos.
- **Administración integrada:** copias de seguridad, permisos de acceso, etc.

Entorno para el despliegue de modelos

Arquitectura en tres niveles



- Distintos requerimientos en función del entorno de producción



La solución de IBM para minería de datos

DB2 Intelligent Miner: Modeling | Scoring | Visualization



- Infraestructura abierta y modular para la consolidación de entornos informacionales y analíticos
- Embeber funcionalidad de minería de datos en otras aplicaciones
 - **DB2 Intelligent Miner Modelling:** construcción y validación de modelos dentro del gestor.
 - **DB2 Intelligent Miner Scoring:** aplicación de modelos bajo demanda y/o en tiempo real.
 - **DB2 Intelligent Miner Visualization:** visualización y análisis interactivo de modelos.
- Funcionalidades soportadas por estándares del mercado
 - Construcción e invocación de modelos mediante SQL/MM
 - Generación de modelos en formato PMML (XML)
 - Permiten la compartición de los modelos generados con cualquier otra aplicación general
- Integración con otras aplicaciones
 - Visualización de modelos dentro de otras aplicaciones que soporten *applets* Java
 - Facilitan la automatización y puesta en producción de los modelos
 - Posibilidad de invocación en tiempo real y/o bajo demanda

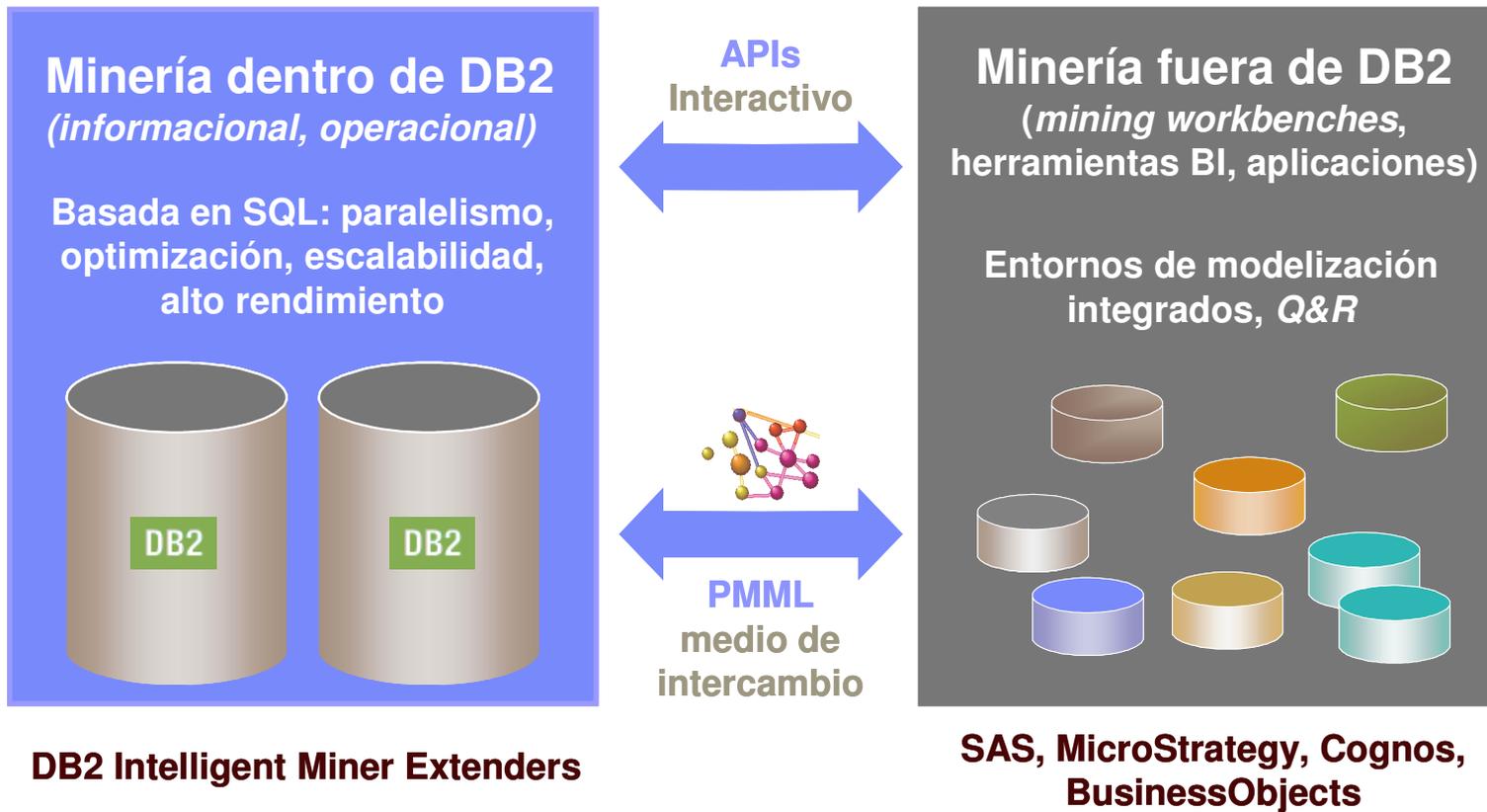


Funciones de minería dentro del gestor

Repositorio central de modelos



- Combinación de paradigmas de modelización en entornos analíticos heterogéneos y distribuidos



Ejemplo: Safeway - Proyecto SmartPad

Sistema de recomendaciones para compras remotas mediante PDA



- Generador de sugerencias como sustituto del mecanismo de compra espontáneo para entornos inalámbricos o electrónicos
- La personalización se lleva a cabo filtrando un conjunto de productos candidatos según la representación del perfil personal del cliente.
- El sistema de recomendaciones funciona de forma autónoma (segmentación de clientes y análisis de asociaciones de productos), aunque permite implementar estrategias de marketing de forma flexible.
- Todo el proceso se lleva a cabo en un servidor centralizado. Las recomendaciones son enviadas al cliente en cada sincronización de las bases de datos de su PDA con el servidor.



Bases de datos personales (PDBs)

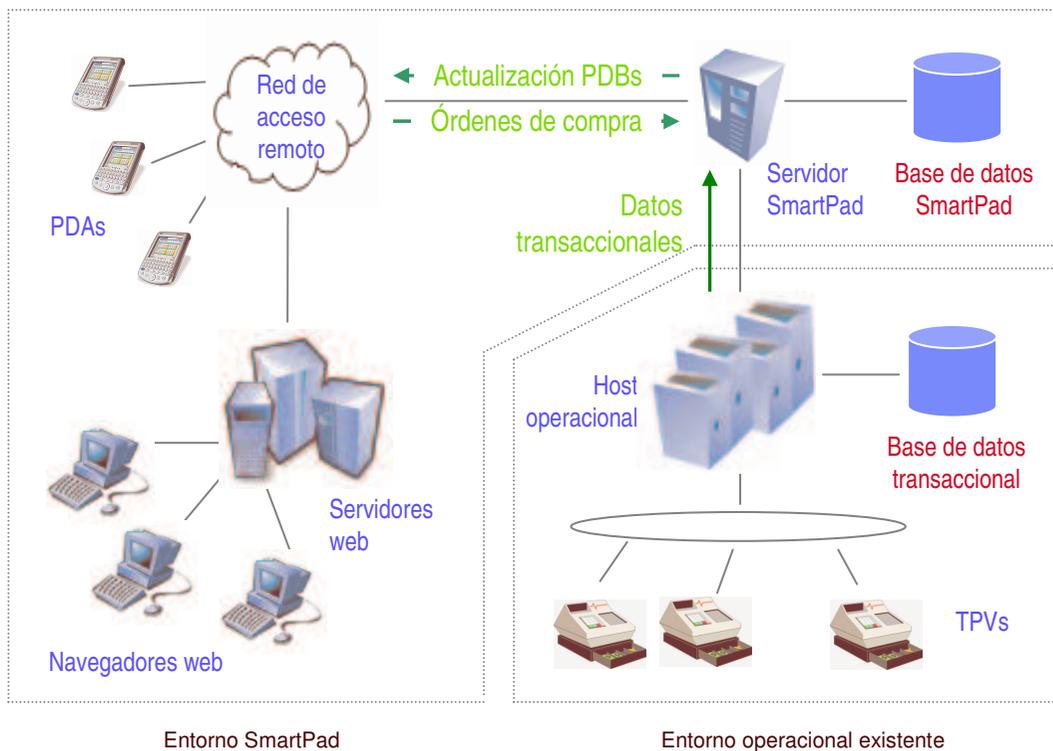


Infraestructura tecnológica

Sincronismo de órdenes de compra y recomendaciones



- Integración del sistema SmartPad en la arquitectura existente



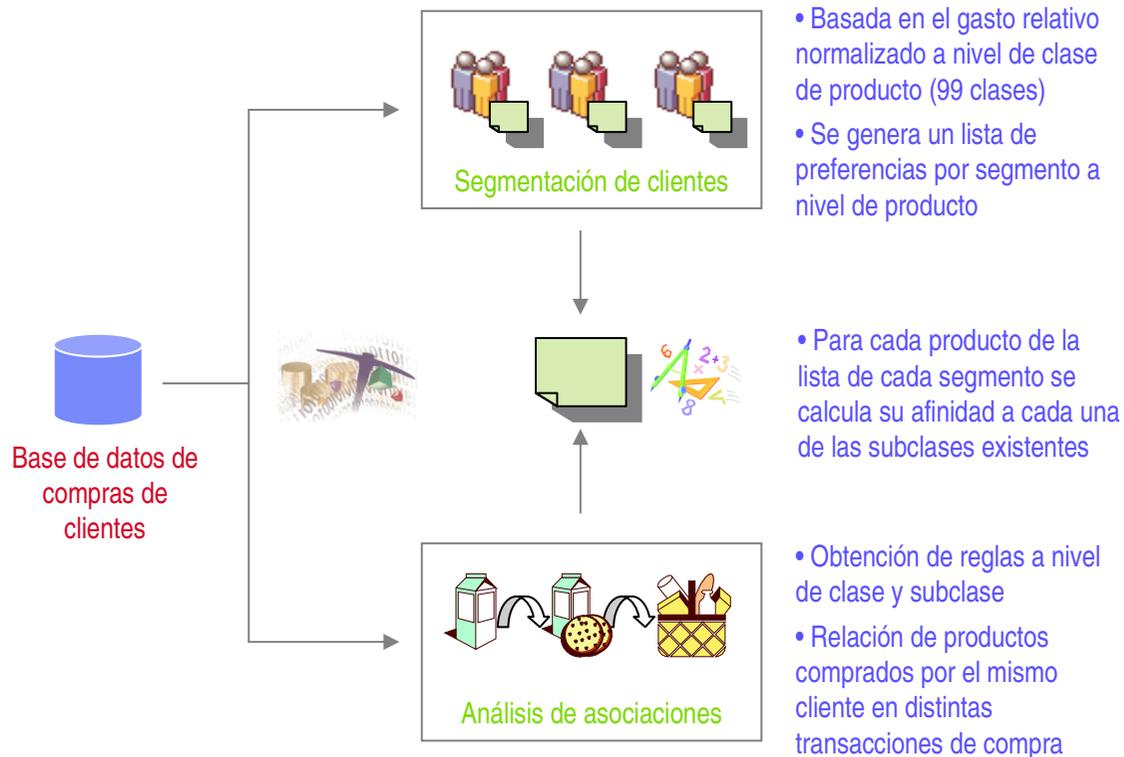
- El servidor operacional envía datos de transacciones al servidor SmartPad.
- La base de datos SmartPad contiene información sobre productos y compras de los clientes.
- El servidor modeliza la información y envía las recomendaciones a los PDAs.

Sistema de recomendaciones

Fase 1. Generación de preferencias y afinidades entre productos



- Segmentación de clientes y detección de asociaciones de productos



- Tanto el modelo de segmentación como el de asociaciones se recalculan de forma automática a intervalos regulares de tiempo.
- Las afinidades producto-subclase se calculan en base a la existencia de reglas de asociación que superen un soporte y confianza mínima.

Sistema de recomendaciones

Fase 2. Selección de los mejores emparejamientos cliente-productos



- Obtención de las recomendaciones en función del perfil del cliente



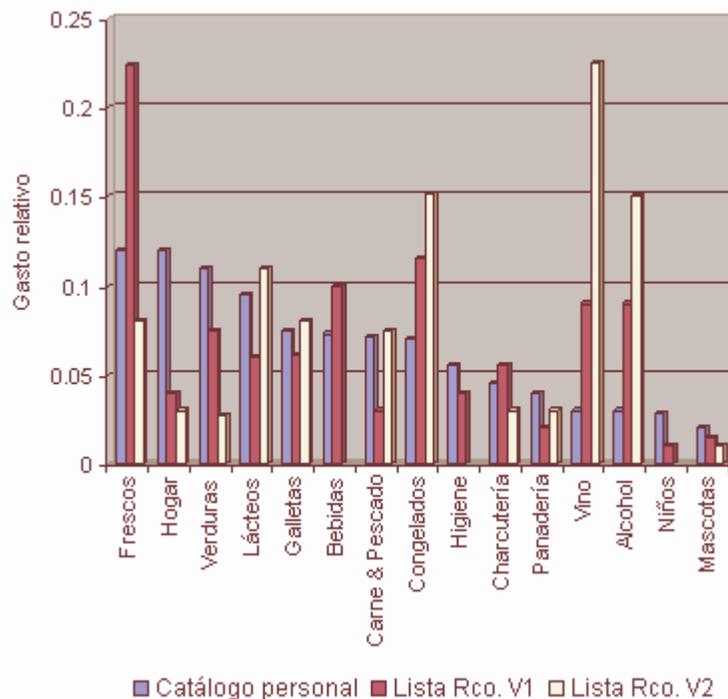
- Se obtiene una lista de productos ordenada según la afinidad esperada (*score*).
- Las afinidades pueden ponderarse según criterios dinámicos (inventario, margen, etc).
- Las recomendaciones finales contienen entre 10 y 20 productos, cubriendo distintas clases y subclases.

Resultados de experiencias piloto

Dos versiones del sistema de recomendaciones



- La afinidad de los productos aceptados es mayor que la de los recomendados en la mayoría de los casos



Versión 1

- Tienda A – 7 meses
- Lista de 10 productos
- Sin ponderar las afinidades obtenidas
- 6.1%** de los pedidos contenían al menos un producto de la lista
- Incremento de ingresos del **0.3%** en los productos recomendados
- Se detectan grupos de productos que suscitan más interés

Versión 2

- Tienda B – 2 meses
- Lista de 20 productos
- Énfasis en referencias nuevas y en productos estrella de la lista
- 25%** de los pedidos con un producto de la lista
- Incremento de ingresos del **2%**, en promedio
- 51%** de los productos correspondían a subclases en las que el cliente hacía 3 meses que no compraba y un **4%** a clases

Fin de la presentación

Víctor López Fandiño
Business Intelligence Solutions Specialist
IBM Software Group

fando@es.ibm.com

