

# **UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA**

## **TECNOLOGÍAS DE LA INFORMACIÓN**



### **EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS**

#### **REPORTE DE SOLUCIÓN DE CASO DE ESTUDIO EN EL QUE INDIQUE LA FORMA EN QUE APLICA LA INTELIGENCIA ARTIFICIAL, MACHINE LEARNING, DATA MINING Y BIG DATA**

***IDGS81N***

**PRESENTA:**

SEBASTIÁN ACOSTA ORTIZ

IAN CARLOS CHÁVEZ ROJO

REGINA CHÁVEZ TAMAYO

IVÁN EDUARDO MARTÍNEZ MARTÍNEZ

ERICK FABIÁN TERRAZAS HERNÁNDEZ

**DOCENTE:**

LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 23 de septiembre de 2025

## Contenido

<b>Introducción al caso .....</b>	3
<b>Justificación del dominio .....</b>	3
<b>Beneficios y KPIs.....</b>	3
<b>Herramientas y lenguajes.....</b>	4
<b>Arquitectura de solución .....</b>	5
<b>    Diagrama .....</b>	6
<b>Conclusiones .....</b>	7
<b>Referencias.....</b>	8

## Introducción al caso

En este proyecto abordamos el caso de RetailMax, un minorista que gestiona millones de compras en tiendas físicas y en línea. Aunque cuenta con una gran cantidad de datos históricos, la empresa aún no los aprovecha plenamente para diferenciar a sus clientes y diseñar estrategias personalizadas.

El reto consiste en desarrollar una solución de segmentación de clientes que permita identificar patrones de compra y características demográficas clave. Con ello se busca crear campañas de marketing personalizado que aumenten la efectividad de las promociones, reduzcan costos y fortalezcan la lealtad de los consumidores.

## Justificación del dominio

Después de analizar las alternativas entre IA, ML, DM y Big Data, concluimos que el dominio más adecuado para este caso es la Minería de Datos (DM).

Nuestra elección se fundamenta en las siguientes razones:

1. La minería de datos está enfocada en descubrir patrones ocultos, asociaciones y estructuras dentro de grandes volúmenes de información.
2. La segmentación de clientes requiere técnicas no supervisadas, en las que no existen etiquetas previas y se necesita clasificar a los usuarios en grupos de manera automática.
3. Las técnicas de clustering, como k-means y el clustering jerárquico, son ampliamente utilizadas en el marketing moderno para identificar perfiles de consumidores con base en su comportamiento.
4. Aunque el Big Data resulta relevante debido al volumen masivo de datos, la verdadera aportación al negocio proviene de aplicar minería de datos para generar conocimiento y estrategias accionables.

En síntesis, la minería de datos se ajusta de manera natural al caso de estudio, ya que su propósito no es solo manejar datos a gran escala, sino convertirlos en información útil para la toma de decisiones de marketing personalizado.

## Beneficios y KPIs

La implementación de esta solución genera beneficios tanto estratégicos como operativos. Entre ellos destacamos:

1. Incremento en la efectividad de campañas publicitarias: al segmentar clientes en grupos homogéneos, se diseñan promociones específicas para cada segmento, lo que se traduce en un aumento proyectado del 20–25 % en la tasa de conversión.
2. Reducción en costos de marketing: al eliminar el enfoque de campañas masivas y sustituirlo por estrategias focalizadas, se optimiza el presupuesto de publicidad, con un ahorro estimado de 15 % en inversión sin afectar los resultados.
3. Mejora en la retención de clientes: gracias a la personalización, se fortalecen las relaciones con los clientes frecuentes y de alto valor, lo que podría traducirse en un incremento del 10–12 % en la recurrencia de compra anual.
4. Mayor comprensión del cliente: los clusters obtenidos permiten identificar perfiles diferenciados (por ejemplo, compradores de alto valor, clientes ocasionales, clientes sensibles a promociones, entre otros), facilitando el diseño de campañas alineadas con las necesidades reales de cada grupo.

Estos beneficios se medirán a través de KPIs específicos como tasa de conversión de campañas, costo por adquisición de cliente (CAC), tasa de retención y valor promedio de compra por segmento.

## Herramientas y lenguajes

Para materializar la propuesta, se seleccionan tecnologías que faciliten el procesamiento, análisis y visualización de la información:

1. **Python con scikit-learn:** se emplea para el cálculo de las métricas RFM, la reducción de dimensionalidad mediante PCA y la implementación de algoritmos de clustering. Su ventaja es que ofrece un ecosistema robusto, probado y con múltiples librerías que aceleran el prototipado.
2. **Apache Spark:** se utiliza para manejar el volumen masivo de datos, garantizando rapidez en el procesamiento distribuido y escalabilidad cuando se integren millones de transacciones históricas.
3. **Tableau o Power BI:** se emplean para la visualización de resultados, permitiendo construir dashboards interactivos en los que los responsables de marketing puedan explorar los segmentos de clientes y tomar decisiones basadas en datos.

De esta manera, cada herramienta cumple un rol complementario: procesamiento (Spark), modelado (Python) y comunicación de resultados (BI dashboards).

## Arquitectura de solución

El flujo de la solución se organiza en etapas que aseguran la preparación, análisis y uso práctico de los datos:

### 1. Fuentes de datos:

- Transacciones históricas de compras en línea y en tienda.
- Información demográfica: edad, género, ubicación geográfica.

### 2. ETL (Extracción, Transformación y Carga):

- Limpieza de datos inconsistentes o incompletos.
- Cálculo de las métricas RFM: recencia, frecuencia y valor monetario por cliente.
- Normalización para estandarizar variables y evitar sesgos.

### 3. Reducción de dimensionalidad (PCA):

- Se aplican componentes principales para capturar al menos el 80 % de la varianza.
- Esto facilita la visualización y evita problemas de sobreajuste.

### 4. Clustering:

- Se implementan dos métodos: k-means y clustering jerárquico.
- El número de clusters se determina mediante el método del codo y el índice de silhouette.

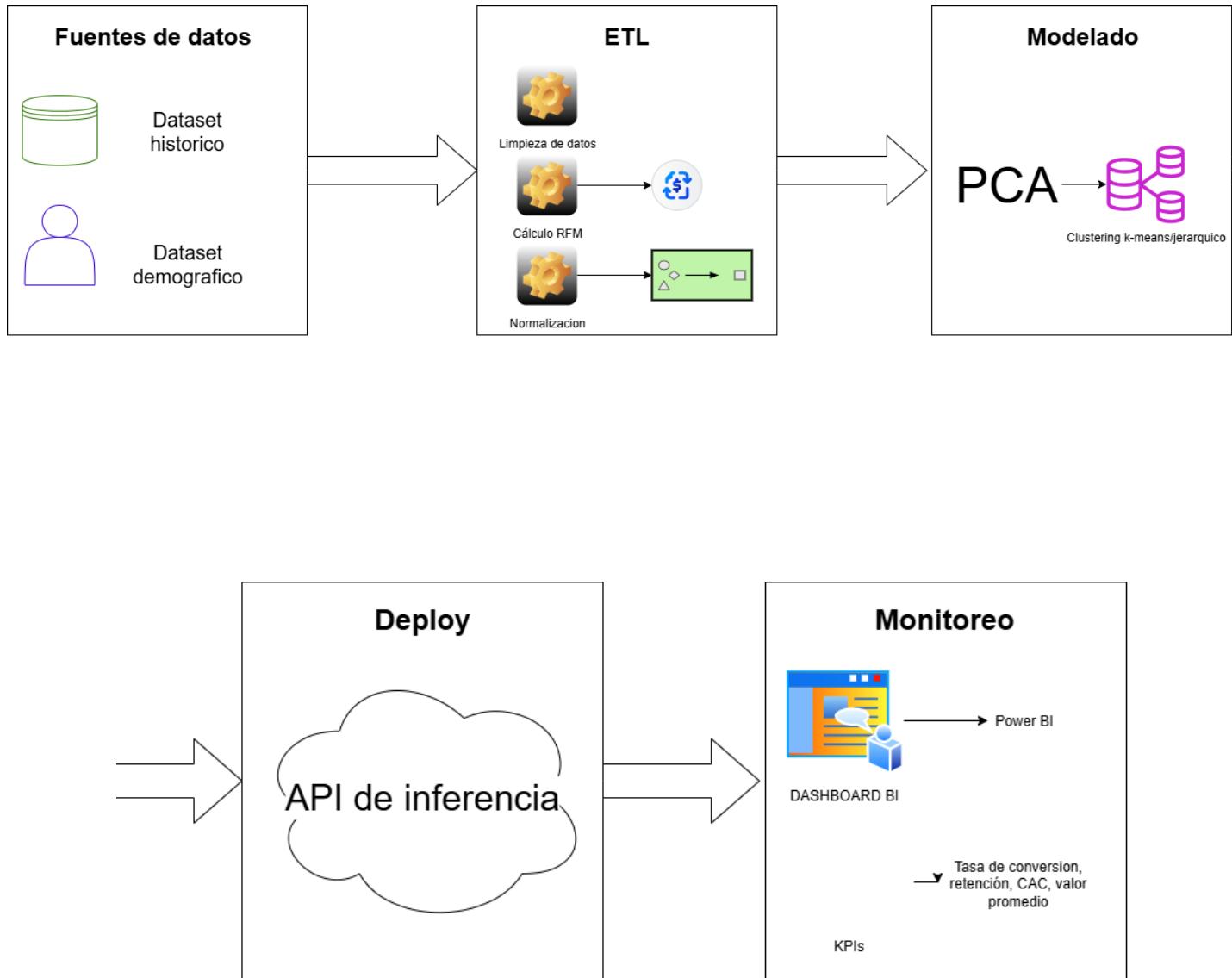
### 5. Perfilado de segmentos:

- Se describen las características más relevantes de cada cluster.
- Se formulan estrategias de marketing específicas para cada grupo identificado.

### 6. Deploy y monitoreo:

- El modelo se expone mediante una API de inferencia.
- Los resultados se integran en un dashboard interactivo, actualizable de forma periódica.

## Diagrama



## Conclusiones

La solución propuesta para RetailMax demuestra que el uso de técnicas de minería de datos aplicadas al análisis de clientes ofrece un valor estratégico significativo. Al transformar datos transaccionales y demográficos en información estructurada mediante métricas como RFM, reducción de dimensionalidad con PCA y algoritmos de clustering, logramos construir un proceso sólido que permite segmentar a los clientes en grupos homogéneos. Esta segmentación facilita la comprensión de los diferentes tipos de consumidores y habilita a la organización para diseñar estrategias de marketing mucho más precisas y efectivas.

Asimismo, la propuesta no solo atiende la necesidad inmediata de agrupar clientes, sino que también establece un marco escalable y adaptable. Al integrar herramientas como Python, Apache Spark y plataformas de visualización, garantizamos que el sistema pueda manejar millones de registros y mantener actualizados los segmentos en función de nuevas transacciones. Esto brinda a RetailMax una ventaja competitiva sostenible, ya que la personalización se convierte en un motor clave para incrementar la conversión de campañas, optimizar costos y fortalecer la relación con los clientes.

Finalmente, este caso nos permitió reflexionar sobre la importancia de aprovechar el potencial de los datos en la toma de decisiones de negocio. El enfoque basado en la minería de datos se convierte en una herramienta poderosa para las organizaciones que buscan innovar en sus estrategias comerciales. Como equipo, consideramos que el siguiente paso será implementar pruebas piloto y extender la solución hacia el uso de modelos predictivos que permitan anticipar comportamientos de compra futuros, consolidando así una estrategia integral de customer analytics que potencie el crecimiento de la empresa.

## Referencias

Adobe Experience Cloud Team. (2023, April 21). *Customer segmentation — targeted marketing made simple*. Adobe. <https://business.adobe.com/blog/basics/define-customer-segmentation>

IBM. (s. f.). *Customer segmentation with data mining*. IBM.  
<https://www.ibm.com/topics/customer-segmentation>

Microsoft. (2024, 3 de mayo). *Introducción a Power BI*. Microsoft Learn.  
<https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>

OpenAI. (2025). Consulta de información sobre segmentación de clientes y minería de datos [Asistente conversacional]. ChatGPT. <https://chat.openai.com/>

Scikit-learn developers. (2024). *Clustering: K-means*. Scikit-learn. <https://scikit-learn.org/stable/modules/clustering.html#k-means>

Silva Ponce de León, C. (2023, 11 de diciembre). *¿Cómo hacer la segmentación de clientes?* Salesforce Blog. <https://www.salesforce.com/mx/blog/segmentacion-de-cliente/Salesforce>