Hackathon 2 – Retail Inteligente

**“Conoce a tu Cliente”**

# Introducción

Las estrategias conocidas como Data-Driven Marketing (Marketing basado en Datos) están tomando un papel primordial en las empresas por varias razones:

* Los datos que se tienen como insumo pueden utilizarse para asignar de manera eficiente los recursos de marketing en las empresas. Lo anterior es muy valorado, ya que evita que se generen campañas que puedan ser costosas y con resultados muy bajos. Adicionalmente se ha visto que las empresas que usan estos métodos de inteligencia de negocios aumentan sus ingresos hasta en un 20%
* El hecho de contar con esta información permite entender mejor a los clientes, entendidas como necesidades y expectativas, para diseñar experiencias comerciales más relevantes y satisfactorias generando de esta forma la fidelización en las marcas.
* Como en cualquier proceso de toma de decisiones, basarse en datos ayuda a predecir en un alto grado comportamientos de consumo por parte de los clientes y de esta manera decidir estrategias tan importantes como un cambio de precio, cambios en la cantidad de producción de artículos, manejo de inventarios, entre otras, lo que resulta más efectivo, esto es hacerlo con base a evidencias y no por intuición.
* Finalmente, mediante la segmentación correcta de los clientes las empresas pueden generar planes de mercadeo que puedan garantizar las ventas futuras de sus productos.

En el área de Data Science y Machine Learning esta problemática tiene un valor importante ya que su solución es aplicable a diferentes sectores como, por ejemplo: Sector Retail, bancario, educación, es decir, para cualquier sector donde el conocimiento del cliente sea fundamental para anticipar su comportamiento y garantizar la sostenibilidad del negocio. Por otro lado, los datos transaccionales y de los clientes son insumos que las empresas del sector tienen al alcance de sus manos para la toma de decisiones asertivas.. Los datos en mención son abundantes y variados, y facilitan las tareas de entrenamiento y evaluación de los modelos.

# Situación problema

MegaShop, cadena nacional líder en retail, quiere potenciar mediante las tecnologías actuales su estrategia comercial. Actualmente, ofrece promociones, descuentos y programas de lealtad sin una segmentación robusta, lo que limita el impacto en ventas y rentabilidad. Esta hackathon busca desarrollar soluciones que permitan conocer mejor al cliente, anticipar sus necesidades y personalizar ofertas mediante el uso de analítica avanzada y aprendizaje de máquina.

Al no contar con una segmentación sólida de sus clientes impide personalizar las estrategias comerciales y maximizar el retorno de inversión en campañas de descuentos y lealtad.

En este sentido, surge la pregunta, cuál es la solución que permita predecir el comportamiento de compra de los clientes en la cadena MegaShop para la toma de decisiones efectiva y asertiva que permita mejorar la rentabilidad del negocio?

# Objetivo General

Desarrollar una solución innovadora, apoyada en el análisis de datos y modelos de aprendizaje de máquina (Machine Learning), que permita segmentar clientes y predecir comportamientos de compra en MegaShop con el fin de personalizar ofertas, anticipar compras y mejorar la rentabilidad del negocio.

# Objetivos Específicos

* Agrupar clientes según patrones de compra mediante técnicas de segmentación como RFM (Recency, Frequency, Monetary Value, entre otros)
* Desarrollar modelos predictivos de monto de compra (regresión) y probabilidad de recompra (clasificación).
* Interpretar los resultados mediante visualizaciones y dashboards claros que apoyen la toma de decisiones de negocio.

# Reto

Con una base de datos rica en información transaccional, el reto consiste en analizar y predecir comportamientos de compra para la toma de decisiones comerciales más efectivas.

# Aportes al reto por categoría

| **Tarea del Reto** | **Responsable Principal** | **Apoyo de otros cursos** |
| --- | --- | --- |
| **Limpieza y exploración de datos (EDA)** | Soluciones IA y Datos | Laboratorio IA y Cloud Computing |
| **Segmentación de clientes (Clustering)** | Soluciones IA y Datos | Laboratorio IA para validación |
| **Modelo predictivo de monto de compra (Regresión)** | Soluciones IA y Datos | Laboratorio IA para ajuste y métricas |
| **Modelo predictivo de recompra (Clasificación)** | Soluciones IA y Datos | Laboratorio IA para interpretabilidad |
| **Métricas, validación cruzada, ajuste fino** | Laboratorio de IA y Cloud Computing | Soluciones IA y Datos |
| **Visualización de insights (gráficos e informes)** | Fábrica de Desarrolladores Web | Soluciones IA (aportes en insights) |
| **Diseño y desarrollo del dashboard interactivo** | Fábrica de Desarrolladores Web | Laboratorio IA para conectividad |
| **Automatización o despliegue en entorno cloud (opcional)** | Laboratorio de IA y Cloud Computing | Fábrica Web para interfaz integrada |
| **Preparación del archivo CSV con predicciones** | Soluciones IA y Datos | — |
| **Preparación de la presentación ejecutiva (pitch)** | Equipo conjunto | Coordinación entre cursos |

# Inspiración y Propósito

Desde el punto de vista del cliente, imagine una experiencia de compra donde cada cliente recibe exactamente lo que necesita, justo cuando lo necesita.

Desde el punto de vista de Megashop, esta puede mejorar el cumplimiento de metas en el ámbito financiero logrando un retorno de inversión y generación de utilidades que garanticen la sostenibilidad de la organización. Así mismo, esto permitirá identificar los productos que satisfagan las necesidades y expectativas de los clientes.

Gracias al análisis avanzado, esta visión puede ser realidad. Con esta hackathon buscamos ayudar a la Cadena de tiendas MegaShop para que comprenda mejor a sus clientes, optimice sus estrategias y mejore la relación con el consumidor.

# Tecnologías Recomendadas

**Lenguaje de programación sugerido:**

* Python: por su ecosistema robusto para análisis de datos y aprendizaje de máquina.

**Librerías útiles:**

* Pandas, NumPy
* Scikit-learn
* Seaborn, Plotly, Matplotlib

**Plataformas recomendadas:**

* Jupyter Notebook, Google Colab, VS Code
* GitHub, Google Drive

**Visualización de resultados (dashboard):**

• Power BI, Tableau, Streamlit, Dash

# Dataset

**Archivo:** megashop\_clients.csv

| Campo | Descripción |
| --- | --- |
| transaction\_id | ID único de la transacción |
| client\_id | ID del cliente |
| transaction\_date | Fecha de compra |
| amount | Valor de la compra en USD |
| store\_location | Ciudad de la tienda |
| product\_category | Categoría del producto |
| payment\_method | Método de pago |
| channel | Canal (Tienda / Web / App) |
| is\_discounted | 1 si aplicó descuento |
| loyalty\_points\_earned | Puntos de lealtad obtenidos |
| previous\_30d\_purchases | Compras en los últimos 30 días |
| client\_tenure\_months | Antigüedad del cliente en meses |

# Entregables Esperados

1. Notebook o script reproducible con todo el flujo de trabajo:

* Exploración de datos (EDA)
* Modelado
* Visualizaciones

2. Presentación de 5 minutos destacando hallazgos clave y acciones recomendadas para MegaShop.

3. Archivo con resultados (segmentos de clientes y predicciones individuales) en formato CSV o Parquet.

# Impacto Esperado

* Tomadores de decisiones más informados gracias a la analítica avanzada.
* Mayor personalización de ofertas y fidelización de clientes.
* Fortalecimiento de habilidades en ciencia de datos, Machine Learning y visualización.
* Aplicación directa del análisis de datos al mundo comercial real.