

RESUMEN EJECUTIVO - PROYECTO AURELION

Retail minorista | IBM & Guayerd | Sprint 4 (Demo 3) - Diciembre 2025

Este documento sintetiza el avance del proyecto, los resultados analíticos y el aporte de los modelos de Machine Learning y el tablero de Power BI. La información se basa en los datasets consolidados, los artefactos de modelo y el reporte BI del proyecto.

Definicion del problema

La direccion requiere convertir datos transaccionales en decisiones concretas para: (1) optimizar inventario y mix de productos, (2) mejorar la retencion de clientes, y (3) priorizar recursos comerciales segun valor esperado. El desafio principal es pasar de datos operativos a analitica confiable y accionable, manteniendo trazabilidad y calidad.

Solucion implementada

La solucion integra BI y ML sobre una base consolidada de ventas minoristas:

- **ETL y normalizacion** de 4 tablas fuente (clientes, productos, ventas, detalle_ventas) con controles de integridad.
- **Dataset master para BI:** `db/base_final_aurelion.csv` con variables transaccionales, temporales y de producto.
- **Dataset para ML:** `db/base_final_ML_clientes.csv` con 67 clientes y 16 features numericas (frecuencia, ticket, mix de categoria, mix de pago, antiguedad, etc.).
- **Power BI:** tablero `Proyecto Aurelion - Power BI.pbix` con KPIs, tendencias, mix de medios de pago y segmentacion por ciudad y categoria.
- **Modelos entrenados:** segmentacion K-Means, churn con Random Forest, y valor de cliente con Random Forest Regressor.

Alcance y calidad de datos

INDICADOR	VALOR	COMENTARIO
Registros transaccionales	343 filas	Base consolidada de ventas
Clientes unicos	67	Universo total de clientes
Transacciones	120	Tickets generados
SKUs	95	Catalogo de productos
Calidad	0 nulos / 0 duplicados	Limpieza completa

Indicadores ejecutivos (negocio)

KPI	RESULTADO	LECTURA EJECUTIVA
Ventas totales	\$2.651.417 ARS	Base suficiente para diagnostico inicial
Ticket promedio	\$7.730 ARS	Nivel medio con dispersion moderada
Ticket mediana	\$6.702 ARS	Menor a la media, indica sesgo por tickets altos
Items por venta	8.5	Ticket volumetrico por compra

Valor promedio por cliente	\$39.573 ARS	Concentracion relevante de cartera
Cliente top	\$132.158 ARS	Oportunidad de fidelizacion focal

Impacto esperado (estimado)

INICIATIVA	IMPACTO ESPERADO	COMO SE MIDE
Retencion focalizada en alto riesgo	Reducir churn y recuperar ventas en cartera critica	% retencion y ventas reactivadas por cohorte
Bundles de limpieza + alimentos	Aumentar ticket y mix de categoria	Variacion del ticket y share de limpieza
Priorizacion por valor esperado	Enfocar recursos comerciales	Incremento de ventas por cliente priorizado
Ajuste de promos por estacionalidad	Capturar picos mensuales	Variacion de ventas en meses pico

Hallazgos principales (Power BI)

Mix de medios de pago

- Efectivo 32.4%, QR 26.5%, tarjeta 20.1%, transferencia 21.0.
- Se observa un peso relevante de metodos digitales (QR + transferencia = 47.5%).

Mix de categorias

- Alimentos 97.1%, limpieza 2.9%.
- La baja participacion de limpieza habilita una estrategia de cross-sell.

Distribucion geografica

CIUDAD	% DE VENTAS
Rio Cuarto	29.9%
Alta Gracia	18.2%
Cordoba	18.2%
Carlos Paz	13.3%
Villa Maria	11.8%
Mendiolaza	8.6%

Estacionalidad

MES	VENTAS (ARS)
-----	--------------

1	\$529.840
2	\$407.041
3	\$388.263
4	\$251.524
5	\$561.832
6	\$512.917

Top productos por facturacion

PRODUCTO	VENTAS (ARS)
Desodorante Aerosol	\$93.800
Queso Rallado 150G	\$89.544
Pizza Congelada Muzzarella	\$85.720
Ron 700Ml	\$81.396
Yerba Mate Suave 1Kg	\$77.560

El top 5 explica 16.1% de la facturacion, senal de catalogo relativamente atomizado.

Resultados de Machine Learning

Segmentacion de clientes (K-Means, K=3)

- **Metricas:** Silhouette 0.1985, Calinski-Harabasz 16.49, Davies-Bouldin 1.65.
- **Distribucion:** 19 / 19 / 29 clientes.

SEGMENTO	PERFIL SINTETICO	INDICADORES CLAVE
Cluster 1	Alto valor y mayor frecuencia	2.74 ventas, \$63.497 ARS total, ticket \$24.468, mix de pago equilibrado
Cluster 0	Ticket alto, baja frecuencia	1.16 ventas, \$41.047 ARS total, ticket \$35.932, 60.5% efectivo
Cluster 2	Bajo valor y baja frecuencia	1.59 ventas, \$22.934 ARS total, ticket \$14.733, mayor uso de tarjeta/transferecia

Resumen operativo de segmentos

SEGMENTO	CLIENTES	FOCO DE ACCION
Cluster 1	19	Mantener fidelizacion y cross-sell
Cluster 0	19	Estimular frecuencia y recompra

Modelo de churn (Random Forest)

- Accuracy: 98.11% train / 57.14% test
- ROC-AUC: 0.57
- Umbral optimo: 0.379
- Clientes de alto riesgo: 20 (29.9% del total) con umbral operativo 0.70

Interpretacion: el modelo detecta patrones, pero exhibe sobreajuste. Es util como semaforo inicial para priorizar acciones de retencion, no como decision automatica.

Modelo de valor del cliente (Random Forest Regressor)

- MAE: \$3.882 (train) / \$7.553 (test)
- RMSE: \$6.870 (train) / \$10.086 (test)
- R2: 0.924 (train) / 0.700 (test)
- Media del target: \$39.573 ARS; mediana: \$34.326 ARS

Interpretacion: el modelo logra una capacidad predictiva razonable para priorizar clientes, con margen de mejora al ampliar historial y enriquecer variables.

Implicancias estrategicas

- **Concentracion geografica:** Rio Cuarto aporta casi un tercio de las ventas; es la plaza critica para retencion y promociones.
- **Oportunidad de mix:** limpieza representa solo 2.9%; se recomienda bundle con alimentos para elevar ticket.
- **Cartera de riesgo:** 20 clientes requieren acciones personalizadas de reactivacion.
- **Segmento premium:** Cluster 1 presenta mejor combinacion de frecuencia y valor.

Riesgos y limitaciones

- Dataset reducido (67 clientes) y rango temporal acotado.
- Segmentacion con calidad moderada (silhouette bajo).
- Churn con sobreajuste; requiere recalibracion y validacion temporal.

Recomendaciones ejecutivas

- Campanas de retencion focalizadas en los 20 clientes de riesgo y en Rio Cuarto.
- Diseñar bundles de limpieza en tickets de alto valor.
- Actualizar el tablero mensualmente y conectar a una fuente actualizable.
- Reentrenar modelos con mayor historial, variables de margen y canal.

Roadmap 30-60-90 dias

HORIZONTE	FOCO	ENTREGABLE
30 dias	Retencion y activacion	Campanas para alto riesgo y clientes top
60 dias	Optimizar mix	Bundles y promociones por categoria
90 dias	Escalar analitica	Reentrenar modelos con mas historia

Glosario breve

- Churn: probabilidad de abandono o inactividad del cliente.
- Silhouette: medida de calidad de clusters (mas alto es mejor).
- R2: que porcentaje de variabilidad explica el modelo.
- MAE/RMSE: error promedio y error cuadratico medio del modelo.

Entregables y trazabilidad

- Power BI: Proyecto Aurelion - Power BI.pbix
- Dataset BI: db/base_final_aurelion.csv
- Dataset ML: db/base_final_ML_clientes.csv
- Modelos y metricas: modelos/*.pkl, modelos/*.json, modelos/*.csv