

# DISEÑO DE POLÍTICAS COMERCIALES PARA UNA EMPRESA DE TECNOLOGÍA PARA SERVICIOS DE SALUD BASADO EN EL COMPORTAMIENTO DE SUS CLIENTES

## INFORME 4

Integrantes: Christopher Lagos  
Profesores: Loreto Martínez G.  
Maximo Bosch P.  
Nicolás Fritis C.

Fecha de entrega: 28 de enero de 2026  
Santiago de Chile

---

## Resumen Ejecutivo

La memoria desarrolla el proyecto “Diseño de políticas comerciales para una empresa de tecnología para servicios de salud basado en el comportamiento de sus clientes”, realizado en Reservo, proveedor chileno de software como servicio para profesionales y centros de salud pequeños y medianos con presencia en Chile, Argentina, Colombia y México. La empresa opera con cerca de 3.500 clientes activos bajo un modelo de suscripción mensual y enfrenta una tasa de bajas cercana a 2,4 por ciento y una gestión de postventa principalmente reactiva. El problema central se define como una caracterización y segmentación conductual insuficiente de la cartera activa, que dificulta priorizar acciones según riesgo de baja y valor económico. El proyecto es desarrollado por el estudiante en modalidad de memoria profesional, en coordinación con el equipo de Customer Success.

El objetivo general es construir un modelo analítico de segmentación conductual que permita gestionar de manera más proactiva las estrategias de retención y expansión. Se sigue la metodología CRISP-DM y se diseña un datamart mensual cliente-mes que integra información de contexto, valor, uso y fricción. Las variables clave incluyen país, rubro, plan comercial, una aproximación al ingreso mensual por cliente, antigüedad, citas recientes, recencia de uso y bloqueos. La etiqueta de salida se define como una baja efectiva en el mes siguiente al corte, lo que plantea un problema de clasificación binaria con horizonte  $T+1$ .

El panel resultante contiene 43.965 observaciones y 1.073 eventos de baja. Sobre esta base se entrena una regresión logística binaria con regularización y ponderación de clases, encapsulada en un flujo reproducible de imputación, tratamiento de extremos, transformaciones y codificación de variables categóricas. Las probabilidades se calibran fuera de muestra y se evalúan en un conjunto de prueba reciente. El modelo alcanza un ROC-AUC cercano a 0,825 y una fuerte capacidad de focalización: el 20 por ciento de clientes con mayor probabilidad estimada concentra alrededor del 70 por ciento de las bajas observadas, con desempeño estable en los segmentos principales.

A partir del datamart y del modelo, el estudiante implementa un prototipo operativo que conecta la analítica con la gestión diaria mediante un script SQL que genera el datamart, un cuaderno en Google Colab que aplica el modelo calibrado y un archivo de Excel que presenta probabilidades, rangos de riesgo y datos de contexto. Se propone una segmentación conductual que combina nivel de riesgo, valor económico aproximado, rubro, plan comercial y motivos de baja, junto con un portafolio inicial de políticas diferenciadas por segmento. La memoria cuantifica una línea base económica, en un escenario de efectividad operativa media por parte del equipo de Customer Success, el modelo permite preservar activos por un monto cercano a los 154 millones de pesos anuales, cifra que valida el impacto financiero del proyecto y justifica plenamente el esfuerzo de implementación. El trabajo concluye que el conjunto formado por el datamart, el modelo de regresión logística calibrado, la segmentación conductual y el prototipo operativo entrega a Reservo una herramienta concreta para priorizar clientes según riesgo, valor y contexto comercial, y deja abiertas líneas de desarrollo en nuevas fuentes de datos, automatización e implementación de pilotos controlados para medir la efectividad y el costo de las acciones de retención y expansión.

# Índice de Contenidos

<b>1. Antecedentes generales</b>	<b>1</b>
1.1. Industria y ambiente competitivo . . . . .	1
1.2. Características de la empresa u organización . . . . .	2
<b>2. Descripción del problema u oportunidad</b>	<b>3</b>
2.1. Antecedentes . . . . .	3
2.1.1. Proceso general . . . . .	3
2.1.2. Etapa de Captación . . . . .	3
2.1.3. Etapa de Ventas . . . . .	3
2.1.4. Etapa de Onboarding . . . . .	3
2.1.5. Etapa de Cliente Activo . . . . .	4
2.1.6. Etapa de Cliente Inactivo . . . . .	4
2.2. Problema encontrado . . . . .	4
2.2.1. Efectos y consecuencias observadas . . . . .	4
2.2.2. Análisis y elección del problema central . . . . .	5
2.2.3. Causas del problema central . . . . .	6
<b>3. Descripción y Justificación del Proyecto</b>	<b>7</b>
3.1. Definición y resultado del proyecto . . . . .	7
3.2. Justificación estratégica . . . . .	7
3.3. Evaluación económica . . . . .	8
3.4. Razones prácticas y técnicas . . . . .	8
3.5. Elección del enfoque . . . . .	9
<b>4. Alcance</b>	<b>10</b>
<b>5. Objetivos del proyecto</b>	<b>11</b>
5.1. Objetivo general . . . . .	11
5.2. Objetivos específicos . . . . .	11
5.3. Indicadores de éxito . . . . .	11
<b>6. Marco Conceptual</b>	<b>12</b>
6.1. Fundamentos de CRM orientado a valor . . . . .	12
6.2. Herramientas de ingeniería y analítica aplicadas . . . . .	12
6.3. Evaluación económica y métricas de desempeño . . . . .	13
6.4. Elección metodológica para el caso de la empresa . . . . .	14
<b>7. Metodología</b>	<b>15</b>
7.1. Fase 1: Comprensión del negocio . . . . .	15
7.2. Fase 2: Comprensión de los datos . . . . .	15
7.3. Fase 3: Preparación de los datos . . . . .	16

7.4.	Fase 4: Modelado . . . . .	16
7.5.	Fase 5: Evaluación . . . . .	17
7.6.	Fase 6: Despliegue y políticas de negocio . . . . .	17
<b>8.</b>	<b>Desarrollo y Resultados</b>	<b>18</b>
8.1.	Fase 1: Comprensión del negocio . . . . .	18
8.1.1.	Contexto de Reservo y modelo de suscripción . . . . .	18
8.1.2.	Problema de gestión asociado a la baja de clientes . . . . .	18
8.1.3.	Definición del objetivo general y de los objetivos específicos . . . . .	18
8.1.4.	Indicadores de éxito y definición de entregables . . . . .	19
8.1.5.	Línea base económica y evaluación basada en el Valor de Vida (CLV) . . . . .	19
8.2.	Fase 2: Comprensión de los datos . . . . .	20
8.2.1.	Inventario de fuentes y tablas principales . . . . .	20
8.2.2.	Estructura y calidad de las tablas . . . . .	21
8.2.3.	Calendario de cortes mensuales . . . . .	21
8.2.4.	Definiciones operativas de estados de cliente . . . . .	21
8.3.	Fase 3: Preparación de los datos . . . . .	21
8.3.1.	Diseño de la estructura del datamart mensual . . . . .	22
8.3.2.	Construcción de variables de contexto, valor, uso y fricción . . . . .	22
8.3.3.	Definición de la etiqueta de baja a horizonte mensual . . . . .	23
8.3.4.	Tratamiento de valores faltantes y valores extremos . . . . .	24
8.4.	Fase 4: Modelado . . . . .	24
8.4.1.	Selección del algoritmo y definición del problema de clasificación . . . . .	25
8.4.2.	Partición temporal y tamaño muestral para entrenamiento y calibración . . . . .	25
8.4.3.	Implementación del flujo de preprocesamiento y del pipeline de modelado . . . . .	25
8.4.4.	Entrenamiento, calibración y elección del modelo final . . . . .	25
8.5.	Fase 5: Evaluación . . . . .	26
8.5.1.	Desempeño global en el conjunto de prueba . . . . .	26
8.5.2.	Calibración de probabilidades y estabilidad por segmentos . . . . .	26
8.5.3.	Pruebas de robustez por país y rubro . . . . .	27
8.5.4.	Conclusiones sobre suficiencia del modelo para uso operativo . . . . .	27
8.6.	Fase 6: Despliegue y políticas de negocio . . . . .	27
8.6.1.	Flujo mensual de scoring sobre la cartera actualizada . . . . .	27
8.6.2.	Artefactos operativos y carpeta de uso para Customer Success . . . . .	28
8.6.3.	Segmentación conductual y políticas operativas . . . . .	28
8.6.4.	Material de apoyo, pruebas en vivo y adopción del prototipo . . . . .	30
<b>9.</b>	<b>Discusiones</b>	<b>32</b>
9.1.	Decisiones de diseño del modelo . . . . .	32
9.2.	Impacto esperado en la operación y rol de Customer Success . . . . .	32
9.3.	Comportamiento del cliente y límites del modelo . . . . .	33

9.4. Limitaciones de datos y riesgos en el uso del prototipo . . . . .	33
<b>10. Conclusiones</b>	<b>34</b>
<b>Referencias</b>	<b>36</b>
<b>Anexo A. Contexto y Problema</b>	<b>38</b>
<b>Anexo B. Indicadores</b>	<b>43</b>
<b>Anexo C. Tablas de desarrollo</b>	<b>44</b>
<b>Anexo D. Código</b>	<b>45</b>
<b>Anexo E. Imágenes de desarrollo</b>	<b>50</b>
<b>Anexo F. Implementación</b>	<b>66</b>

## Índice de Figuras

1. Flujo General del Cliente por Área. Fuente: elaboración propia . . . . .	3
2. Churn 3 años (ene-2023 a jul-2025) y tendencia. Fuente: elaboración propia .	5
3. Número de clientes al inicio de cada mes (ene-2023 a jul-2025). Fuente: elaboración propia . . . . .	5
A.1. Organigrama de Reservo, Agosto 2025. Fuente: elaboración propia . . . . .	38
A.2. Etapa de Adquisición en Reservo. Fuente: elaboración propia . . . . .	38
A.3. Etapa de Venta en Reservo. Fuente: elaboración propia . . . . .	39
A.4. Etapa de Onboarding del cliente en Reservo. Fuente: elaboración propia . . .	40
A.5. Etapa de cliente activo en Reservo. Fuente: elaboración propia . . . . .	41
A.6. Etapa de Cliente inactivo en Reservo. Fuente: elaboración propia . . . . .	42
E.1. Extracto código SQL construcción de datamart. Fuente: elaboración propia .	50
E.2. Extracto de código modelo Python/Colab. Fuente: elaboración propia . . . .	50
E.3. Porcentaje de nulos por variable. Fuente: elaboración propia . . . . .	51
E.4. Tipos de datos de las variables. Fuente: elaboración propia . . . . .	51
E.5. Estadísticas descriptivas de variables numéricas. Fuente: elaboración propia .	52
E.6. Porcentaje de clientes por rubro. Fuente: elaboración propia . . . . .	52
E.7. Porcentaje de clientes por país. Fuente: elaboración propia . . . . .	52
E.8. Porcentaje de clientes por plan. Fuente: elaboración propia . . . . .	53
E.9. Proporción de planes por rubro. Fuente: elaboración propia . . . . .	53
E.10. Proporción de rubros por plan. Fuente: elaboración propia . . . . .	54
E.11. Comparativa de proporciones de activos vs bloqueados por rubro. Fuente: elaboración propia . . . . .	54
E.12. Comparativa de proporciones de activos vs bloqueados por plan. Fuente: elaboración propia . . . . .	55

E.13.	Distribución y boxplot de <i>tenure_meses</i> . Fuente: elaboración propia . . . . .	55
E.14.	Distribución y boxplot de <i>arpa_proxy</i> . Fuente: elaboración propia . . . . .	55
E.15.	Distribución y boxplot de <i>bloqueos_90d</i> . Fuente: elaboración propia . . . . .	56
E.16.	Distribución y boxplot de <i>citas_30d</i> . Fuente: elaboración propia . . . . .	56
E.17.	Distribución y boxplot de <i>recencia_uso_dias</i> . Fuente: elaboración propia . . . . .	56
E.18.	Matriz de correlaciones entre variables numéricas. Fuente: elaboración propia . . . . .	57
E.19.	Diagrama de Datamart construido. Fuente: elaboración propia . . . . .	57
E.20.	Resultado python capacidad de discriminación del modelo. Fuente: elaboración propia . . . . .	58
E.21.	Resultado python captura acumulada. Fuente: elaboración propia . . . . .	58
E.22.	Resultado python métricas operativas. Fuente: elaboración propia . . . . .	58
E.23.	Resultado python calibración global. Fuente: elaboración propia . . . . .	59
E.24.	Resultado python calibración por país. Fuente: elaboración propia . . . . .	59
E.25.	Resultado python calibración por plan. Fuente: elaboración propia . . . . .	60
E.26.	Resultado python calibración por rubro. Fuente: elaboración propia . . . . .	60
E.27.	Gráfico de curva ROC. Fuente: elaboración propia . . . . .	60
E.28.	Gráfico curva de calibración. Fuente: elaboración propia . . . . .	61
E.29.	Gráfico curva de captura acumulada. Fuente: elaboración propia . . . . .	61
E.30.	Gráfico curva precisión-recall. Fuente: elaboración propia . . . . .	62
E.31.	Gráfico curva ROC Chile-México-Colombia. Fuente: elaboración propia . . . . .	62
E.32.	Gráfico curva ROC Chile + Medicina. Fuente: elaboración propia . . . . .	63
E.33.	Gráfico curva ROC medicina. Fuente: elaboración propia . . . . .	63
E.34.	Gráfico curva ROC México. Fuente: elaboración propia . . . . .	64
E.35.	Gráfico curva ROC Colombia. Fuente: elaboración propia . . . . .	64
E.36.	Gráfico curva captura acumulada Colombia. Fuente: elaboración propia . . . . .	65
E.37.	Gráfico curva captura acumulada México. Fuente: elaboración propia . . . . .	65
F.1.	Colab de uso Customer Success. Fuente: elaboración propia . . . . .	66
F.2.	Excel de salida del modelo. Fuente: elaboración propia . . . . .	66
F.3.	Carpetas entregadas a Customer Success. Fuente: elaboración propia . . . . .	67
F.4.	Página de Notion elaborada para Customer Success. Fuente: elaboración propia . . . . .	67
F.5.	Extracto guía de implementación y uso elaborada para Customer Success. Fuente: elaboración propia . . . . .	68
F.6.	Tutorial en video elaborado para Customer Success. Fuente: elaboración propia . . . . .	69

## Índice de Tablas

1.	Valor Patrimonial Retenido (CLV acumulado) según escenarios de efectividad operativa. . . . .	20
B.1.	Métricas asociadas a clientes. Fuente: Elaboración propia . . . . .	43
C.1.	Bloqueados y actividad de citas por rubro. Fuente: elaboración propia . . . . .	44
C.2.	Bloqueados y actividad de citas por plan. Fuente: elaboración propia . . . . .	44
C.3.	Promedio de mensualidad por rubro (CLP). Fuente: elaboración propia . . . . .	44

# 1. Antecedentes generales

## 1.1. Industria y ambiente competitivo

El sector de salud en Chile y América Latina vive un proceso sostenido de digitalización. El Observatorio de Desarrollo Digital de la CEPAL muestra que muchas empresas, en especial MIPYMES, siguen teniendo una presencia en línea limitada, lo que refuerza la necesidad de soluciones que modernicen la gestión de servicios de salud (CEPAL, 2024). Al mismo tiempo, el ecosistema regional de innovación registra más de un millar de empresas de salud digital, con foco en prevención, gestión y soporte a la toma de decisiones (IDB Lab and HolonIQ, 2024).

En Chile este dinamismo se observa en programas como Start-Up Chile, que desde 2010 ha apoyado a más de 2,200 startups con un portafolio valorizado en miles de millones de dólares (Paredes, 2025). La inversión de capital de riesgo en *healthtech* latinoamericano también ha crecido, superando los mil millones de dólares en 2021 (Munar, 2023). Siguiendo la clasificación de la OMS, la salud digital incluye soluciones para pacientes, profesionales y gestores del sistema, entre ellas agendamiento, teleconsulta y analítica (World Health Organization, 2018).

La pandemia aceleró estos procesos. La expansión de la telemedicina y de modelos híbridos de atención elevó las expectativas de pacientes y equipos de salud en torno a continuidad de cuidados, acceso remoto y coordinación entre prestadores (Superintendencia de Salud de Chile, 2025). Persisten, sin embargo, brechas de adopción y recursos que limitan el uso sostenido de estas herramientas, en especial en prestadores pequeños y medianos (Revista Chilena de Salud Pública (Universidad de Chile), 2025). En este contexto surge un ecosistema de software competitivo, donde se distinguen tres tipos de actores:

- **Startups y empresas nativas digitales**, ágiles e innovadoras, enfocadas en nichos específicos, pero con desafíos para escalar y generar confianza de largo plazo.
- **Empresas locales consolidadas**, con conocimiento del marco regulatorio y cultural chileno y bases de clientes estables, aunque con procesos de innovación más lentos.
- **Actores internacionales**, con tecnologías maduras y marcas globales, que deben adaptar sus soluciones a las particularidades normativas del país, como los requisitos de seguridad y privacidad establecidos en la Ley de Derechos y Deberes del Paciente (Biblioteca del Congreso Nacional de Chile, 2012).

El mercado direccionable para Reservo combina profesionales individuales y centros ambulatorios privados. A marzo de 2024 el registro de la Superintendencia de Salud contabiliza 433.431 profesionales habilitados, base potencial para cuentas individuales y consultas privadas (Clínicas de Chile A.G., 2025). En el segmento institucional, a febrero de 2025 existían 555 prestadores privados acreditados, de los cuales 175 corresponden a centros de atención abierta con demanda directa por agendamiento y gestión de pacientes (Superintendencia de Salud (Gobierno de Chile), 2025).

El espacio competitivo local ya muestra escala. AgendaPro reporta más de 100.000 profesionales y decenas de miles de negocios gestionados, con más de 100 millones de citas procesadas (BusinessWire, 2025). Medilink (Healthatom) declara miles de clientes y decenas de millones de citas acumuladas (Healthatom (Medilink), 2025). Encuadrado, enfocada en profesionales independientes, trabaja con cientos de centros y más de 5.000 profesionales activos (Encuadrado, 2025). Reserva compete en este entorno, con foco en salud ambulatoria y prestadores pequeños y medianos.

## 1.2. Características de la empresa u organización

RESERVO es una empresa chilena de tecnología SaaS que actúa como socio tecnológico para profesionales y centros de salud pequeños y medianos en Chile, Argentina, Colombia y México (RESERVO, 2025). Su modelo se orienta a un segmento que suele quedar fuera de la oferta de grandes proveedores orientados a hospitales y grandes cadenas. La base actual de clientes se compone de especialistas independientes, dentistas, kinesiólogos, psicólogos y clínicas medianas en contextos ambulatorios.

La empresa ofrece cuatro planes (Individual, Básico, Reserva y Enterprise) que se diferencian por funcionalidades, número de usuarios y nivel de soporte. Todos incluyen agenda y reservas en línea con recordatorios, ficha clínica electrónica, reportes y un módulo financiero. Los planes superiores agregan control de permisos, automatizaciones, centros y usuarios ilimitados y soporte prioritario. La plataforma se integra con medios de pago, emisión de documentos tributarios, conectores de marketing y redes sociales, comercio electrónico, receta electrónica, servicios de exámenes a domicilio e integraciones contables.<sup>1</sup>

El problema que RESERVO busca resolver es que muchos profesionales destinan una parte importante de su tiempo a tareas administrativas (gestión de agendas, recordatorios, cobros) y cuentan con pocas herramientas para competir digitalmente y atraer pacientes.<sup>2</sup> La propuesta de valor de la plataforma se resume en tres ejes: eficiencia operativa, gestión financiera y marketing digital, ofreciendo una solución accesible y escalable que mejora eficiencia, rentabilidad y presencia en línea en un mercado competitivo.

La organización se estructura de forma jerárquica bajo la dirección del CEO, con liderazgos en Post-venta, Customer Success, Comercial, Marketing y Desarrollo. Esta configuración combina capacidades técnicas, comerciales y de acompañamiento al cliente, y sostiene el crecimiento de la plataforma con equipos ajustados al tamaño actual de la empresa.

---

<sup>1</sup> Descripción basada en la información pública disponible en la sección “Características” del sitio web de RESERVO. <https://www.reservo.cl/caracteristicas>

<sup>2</sup> Análisis basado en una entrevista con la Customer Success Manager de RESERVO, realizada el 15 de agosto de 2025.



## 2. Descripción del problema u oportunidad

### 2.1. Antecedentes

#### 2.1.1. Proceso general

La Figura 1 muestra el flujo general del cliente a lo largo de su ciclo de vida en Reservo, desde el primer contacto con Marketing hasta su permanencia o pérdida. El recorrido incluye las etapas de captación y venta (SDR y ejecutivos comerciales), incorporación a la plataforma (Onboarding a cargo de Customer Success), atención de Soporte y gestión de Retención, donde se define si el cliente se mantiene activo o se produce churn.

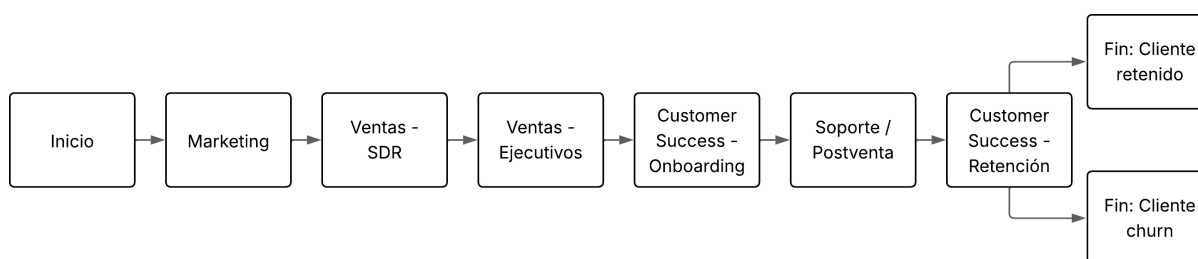


Figura 1: Flujo General del Cliente por Área. Fuente: elaboración propia

#### 2.1.2. Etapa de Captación

En la etapa de captación, ilustrada en el anexo A.2, Marketing ejecuta campañas que transforman a un visitante desconocido en un **lead** registrado. El equipo de **Ventas SDR** contacta a ese lead, valida que cumpla el perfil objetivo y, si la calificación es positiva, lo convierte en **lead calificado**. Finalmente, el SDR agenda una demostración con un ejecutivo de ventas y envía una cotización, momento en que el contacto pasa a estado de **Oportunidad de venta** y queda listo para la etapa comercial.

#### 2.1.3. Etapa de Ventas

El proceso de ventas, descrito en el anexo A.3, comienza cuando la oportunidad asiste a la demo del producto. Si decide contratar, el ejecutivo configura la cuenta en el sistema y el estado cambia a **precliente**. Tras el pago inicial, el registro pasa a **cliente** y se traspasa formalmente al equipo de **Customer Success**, que inicia el Onboarding y cierra la etapa comercial.

#### 2.1.4. Etapa de Onboarding

En el Onboarding, detallado en el anexo A.4, Customer Success prepara la cuenta, agenda una primera capacitación y monitorea el uso inicial. Según este comportamiento, el cliente puede ser clasificado como de **riesgo alto** o **riesgo medio**, lo que dispara contactos proactivos para asegurar la adopción. Si no se observan alertas en un periodo definido, el Onboarding se considera exitoso y el cliente pasa al uso regular de la plataforma.

### 2.1.5. Etapa de Cliente Activo

La etapa de **cliente activo**, ilustrada en el anexo A.5, se centra en retención y expansión. Customer Success monitorea indicadores de salud de la cuenta y activa acciones de seguimiento cuando detecta bajo uso u otros signos de riesgo. En casos sanos se buscan oportunidades de *upselling*; si el cliente acepta la oferta y paga, se actualiza su plan y se refuerza su condición de **cliente vigente**. En paralelo se gestionan tickets de soporte y procesos de facturación, donde problemas de pago recurrentes pueden terminar en la pérdida del cliente.

### 2.1.6. Etapa de Cliente Inactivo

El flujo de **cliente inactivo**, descrito en el anexo A.6, comienza cuando un cliente solicita cancelar el servicio. Customer Success procesa la baja, registra el motivo y formaliza el estado de **dado de baja**. El diagrama contempla también la reactivación: si un cliente decide volver, se gestiona un nuevo contrato y se reactiva su cuenta, devolviéndolo al estado de **cliente activo** dentro del ciclo general.

## 2.2. Problema encontrado

Para sostener su crecimiento en el mercado SaaS, Reservo necesita extraer más valor de su base de clientes. Hoy enfrenta dificultades para retener y expandir cuentas de forma sistemática, lo que se traduce en pérdidas de ingresos, crecimiento moderado y uso poco eficiente de los recursos de postventa.

### 2.2.1. Efectos y consecuencias observadas

Los resultados actuales de la gestión del ciclo de vida del cliente se reflejan en tres efectos principales:

- **Pérdida recurrente de clientes e ingresos.** La empresa presenta una **tasa de churn** mensual de **2,46 %**, equivalente a **70 bajas** en lo que va de 2025<sup>3</sup>. Desde enero de 2022 el churn promedia **2,77 %**. La Figura 2 muestra su evolución junto con el número de clientes al inicio de cada mes (Figura 3). Aunque la serie presenta una ligera tendencia a la baja, el indicador se mantiene desde hace casi dos años en el rango 2–3 % sin perforar el 2 %. Esto es relevante porque los benchmarks para empresas SaaS recomiendan apuntar a menos de 2 % mensual para alinearse con las mejores prácticas del mercado (Jain, 2023).
- **Expansión y reactivación no sistemáticas.** La empresa genera valor adicional, pero sin un crecimiento acelerado. En 2024 se registraron **57 upsellings** y en 2025 se han concretado **60** hasta la fecha. Respecto de la reactivación, en 2024 se recuperaron **100 clientes** y en 2025 van **89**<sup>4</sup>. Las cifras indican un avance moderado en expansión y

<sup>3</sup> La **Tasa de Churn** se define en la tabla B.1. El valor y el número de bajas corresponden a datos internos para 2025. Fuente: Base de datos de Reservo.

<sup>4</sup> Datos de upsellings y reactivaciones según registros internos. Fuente: Base de datos de Reservo.

una posible pérdida de fuerza en reactivación, lo que sugiere ausencia de una estrategia proactiva y repetible.

- **Asignación ineficiente de recursos de postventa.** El equipo de Customer Success suele intervenir tarde y con ofertas poco ajustadas al perfil del cliente en riesgo. No existe una clasificación operativa de riesgo durante la etapa activa (más allá del Onboarding), por lo que muchas acciones se disparan cuando la probabilidad de abandono ya es alta. Se termina priorizando la contención de clientes al borde del churn y se desaprovechan intervenciones tempranas con mayor potencial de retención o expansión.

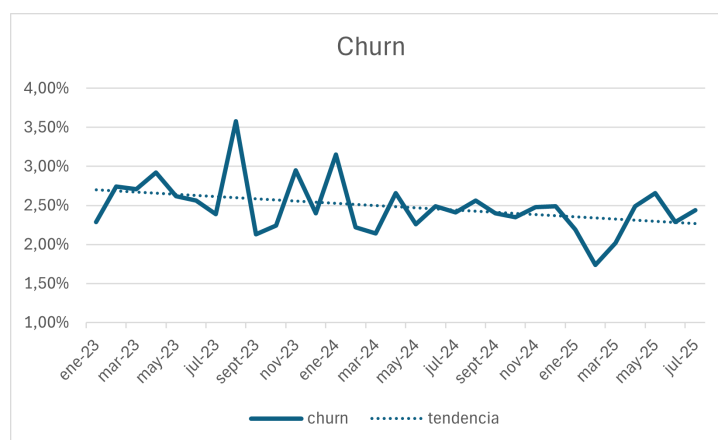


Figura 2: Churn 3 años (ene-2023 a jul-2025) y tendencia. Fuente: elaboración propia

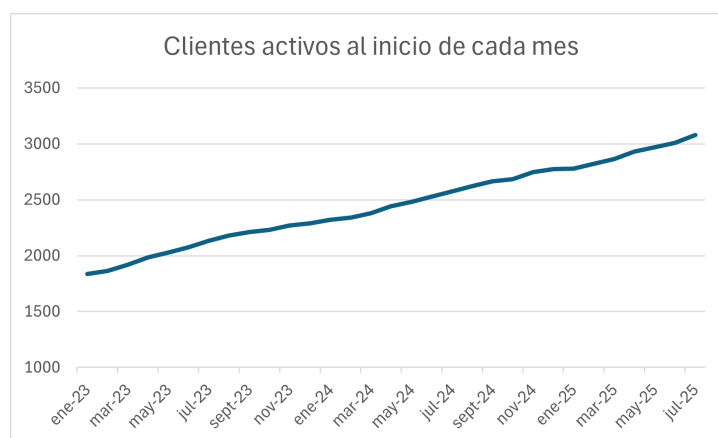


Figura 3: Número de clientes al inicio de cada mes (ene-2023 a jul-2025). Fuente: elaboración propia

### 2.2.2. Análisis y elección del problema central

Para interpretar estos efectos se utiliza como marco la **Gestión del Ciclo de Vida del Cliente (Customer Lifecycle Management)**, que plantea que las acciones deben

adaptarse a la etapa y comportamiento de cada cuenta. Desde esta perspectiva, la tasa de abandono elevada, el crecimiento limitado del valor y la baja efectividad de postventa no son causas de fondo, sino consecuencias de una falla operativa más profunda. Se busca entonces el problema que, al abordarse, tenga mayor impacto sobre los demás síntomas. El problema central seleccionado es:

***Insuficiente caracterización y segmentación conductual de la cartera activa en postventa***

**2.2.3. Causas del problema central**

El origen del problema no está en la falta de datos, sino en cómo se utilizan para orientar la gestión. Entre las causas principales se identifican:

- **Decisiones basadas en indicadores de resultado.** Las estrategias se activan a partir de métricas retrospectivas, como la **tasa de churn** o un **NPS** de 52<sup>5</sup>. Estos indicadores describen lo que ya ocurrió, pero no permiten diferenciar el tratamiento entre clientes antes de que el riesgo se materialice.
- **Análisis estático de la información.** Se observa el estado actual de cada cliente, pero no su trayectoria en el tiempo. La ausencia de una mirada longitudinal dificulta detectar señales tempranas, como una baja gradual en el uso de la plataforma.
- **Falta de criterios conductuales para segmentar.** No se han definido patrones de uso que distingan clientes saludables, en riesgo o con potencial de expansión. Sin estos criterios, la segmentación sigue siendo principalmente demográfica o comercial y no captura el comportamiento dinámico.
- **Enfoque operativo reactivo.** El trabajo de Customer Success se organiza para responder a eventos negativos (quejas, impagos, bloqueos) más que para anticiparse al riesgo. No existe un esquema sistemático que combine datos de comportamiento y valor económico para priorizar acciones personalizadas sobre la cartera activa.

---

<sup>5</sup> El **NPS** se define en la tabla B.1. El valor corresponde al último resultado medido por la empresa. Fuente: Base de datos de Reservo.

## 3. Descripción y Justificación del Proyecto

### 3.1. Definición y resultado del proyecto

El proyecto titulado “Diseño de políticas comerciales para una empresa de tecnología para servicios de salud basado en el comportamiento de sus clientes” consiste en el diseño de un modelo analítico y de un conjunto de herramientas que apoyan la gestión comercial de la cartera de clientes de Reservo. El foco está en reducir la baja de clientes y en aprovechar mejor las oportunidades de expansión comercial, tanto en la etapa de onboarding como cuando el cliente ya se encuentra en la fase de uso normal del servicio.

El eje central es un modelo que estima, para cada cliente, la probabilidad de que deje de usar la plataforma en un horizonte mensual. Este cálculo se realiza de manera recurrente a partir de información histórica de comportamiento y facturación, de modo que la empresa cuente con una señal temprana de riesgo desde el inicio de la relación y durante toda la vida activa del cliente. Sobre la base de estas probabilidades se construyen listados y segmentaciones de la cartera que permiten identificar a los clientes con mayor probabilidad de abandono, priorizar esfuerzos de retención y apoyar decisiones de expansión en los grupos con mayor potencial, orientando así los recursos hacia el mayor impacto esperado en ingresos futuros.

Para hacer operable este enfoque, el proyecto entrega varios artefactos concretos. Se construye un datamart que reúne y organiza la información relevante de clientes y uso del sistema, se entrena y documenta el modelo de riesgo de baja y se implementa un flujo simple de uso en un notebook de Google Colab y en archivos de Excel, que permite al equipo de Customer Success aplicar el modelo sobre bases mensuales actualizadas sin necesidad de conocimientos avanzados de programación. Cada ciclo de uso genera archivos con la lista de clientes, su nivel de riesgo y la información necesaria para planificar acciones. A partir de estos insumos se definen políticas comerciales prácticas, entendidas como reglas y recomendaciones de acción diferenciadas por tipo de cliente y nivel de riesgo, de modo que el proyecto no solo describe un modelo, sino que entrega una forma concreta de usarlo en la gestión cotidiana de la empresa.

### 3.2. Justificación estratégica

La realización de este proyecto se justifica por su alineación directa con la necesidad de la empresa de optimizar la rentabilidad de su cartera y asegurar un crecimiento sostenible. La gestión proactiva del cliente basada en datos es un elemento clave para la competitividad y la rentabilidad en mercados digitales (Kumar and Reinartz, 2018).

Las razones estratégicas se basan en tres pilares principales:

- Mitigar la pérdida de ingresos por bajas de clientes. La tasa mensual de bajas se estima en torno a 2,46 % y afecta de manera directa los ingresos recurrentes. La evolución histórica se muestra en la Figura 2. Un modelo predictivo permite anticipar señales de deserción y entregar al equipo comercial una lista de clientes con mayor riesgo, de modo que se puedan activar acciones de retención más oportunas, en línea con la literatura de ciclo de vida del cliente (Lemmens and Croux, 2008).

- Capitalizar oportunidades de expansión. Aumentar el valor económico de cada cliente mediante ventas adicionales y aumento de plan permite aprovechar mejor la base instalada (Kumar and Reinartz, 2018). La evolución observada en los indicadores internos (57 en 2024 frente a 60 en 2025) sugiere que existe espacio para sistematizar estas oportunidades y convertirlas en una fuente más estable de crecimiento.
- Optimizar el esfuerzo de Customer Success. Un trato homogéneo de toda la cartera genera sobrecostos y baja efectividad. La segmentación basada en datos permite distinguir clientes de mayor riesgo o mayor potencial y asignar prioridades claras de trabajo. Esto facilita organizar la agenda del equipo, concentrar los esfuerzos donde el impacto esperado es mayor y se alinea con prácticas de gestión de relaciones con clientes (Kumar and Reinartz, 2018).

### 3.3. Evaluación económica

La operación actual de Reservo cuenta con una cartera aproximada de 3.500 clientes activos. Estos generan un ingreso mensual recurrente cercano a los \$205 millones, lo que proyecta una facturación anual en torno a los \$2.400 millones de pesos. El pago promedio por cliente (ARPA) se sitúa en \$58.700 mensuales.

Históricamente, la tasa de cancelación (churn) es del 2,4 % mensual, lo que implica la pérdida de más de 80 clientes cada mes. Desde una perspectiva financiera, la salida de un cliente no representa solo la pérdida de su mensualidad, sino la destrucción de un activo de largo plazo. Se estima que cada cliente que abandona la compañía representa una pérdida de valor patrimonial cercana a \$1,28 millones, considerando los flujos futuros que deja de aportar.

El proyecto implementa un modelo predictivo para frenar esta pérdida de valor. Al identificar preventivamente a los clientes en riesgo, el equipo de *Customer Success* puede ejecutar acciones de retención focalizadas. Las proyecciones indican que, en un escenario de efectividad media, esta gestión permite rescatar el valor de 120 clientes al año. Esto se traduce en la preservación de activos por un monto aproximado de \$154 millones anuales. Esta cifra demuestra que la retención de clientes tiene un impacto financiero significativo y directo sobre la rentabilidad de la empresa, complementando las estrategias de adquisición de nuevas cuentas.

### 3.4. Razones prácticas y técnicas

El enfoque elegido se apoya en las capacidades y recursos que ya existen en Reservo. La empresa cuenta con bases de datos históricas de uso de la plataforma, estados de cliente, planes, facturación y tickets de soporte, además de una infraestructura básica para analítica y herramientas de BI. Sobre esta base es posible construir un flujo reproducible que incluye extracción de datos, depuración, construcción de un datamart, segmentación de clientes y definición de reglas de acción conectadas con la operación diaria.

La segmentación conductual permite definir riesgos y oportunidades a partir de variables observables y actualizables en el tiempo, como patrones de uso y comportamiento de pago.

Los modelos utilizados privilegian interpretabilidad, trazabilidad y control de supuestos, de modo que sus resultados puedan ser revisados y entendidos por los equipos de negocio, evitando que el modelo se perciba como una caja negra y favoreciendo su integración al lenguaje cotidiano de la empresa. A lo largo del desarrollo del proyecto se prepara además material práctico para el uso del modelo: una carpeta estructurada con los scripts de generación del datamart, el notebook de scoring en Google Colab y los archivos de Excel que presentan los resultados en formatos conocidos para Customer Success, junto con tutoriales escritos y sesiones de capacitación. El resultado esperado es un diseño validado en datos históricos y un conjunto de políticas comerciales que el equipo puede aplicar de manera autónoma, en línea con la gestión de clientes basada en valor y con el uso de analítica para priorizar esfuerzos de retención y expansión (Kumar and Reinartz, 2018).

### **3.5. Elección del enfoque**

Se elige un modelo analítico de segmentación conductual con scoring y reglas de acción. La motivación principal es que permite identificar grupos de clientes según su comportamiento y asignar a cada grupo un nivel de prioridad comercial concreto. Esto facilita orientar los esfuerzos del equipo hacia quienes presentan mayor probabilidad de baja o mayor potencial de crecimiento, utilizando la información disponible en la base de datos de la empresa.

Frente a alternativas centradas solo en reportes descriptivos, este enfoque agrega una capa de predicción y ordenamiento que entrega listas accionables. Frente a reglas puramente heurísticas, aporta criterios contruidos a partir de datos históricos y actualizables en el tiempo. Al mismo tiempo mantiene una estructura sencilla: variables comprensibles para el negocio, reglas claras de interpretación y resultados que se pueden revisar en herramientas habituales como hojas de cálculo. De este modo, el enfoque combina impacto esperado en retención y expansión, interpretabilidad para los equipos comerciales y compatibilidad con la infraestructura actual, en línea con la literatura de CRM y gestión del ciclo de vida del cliente (Kumar and Reinartz, 2018).

## 4. Alcance

El alcance del proyecto abarca el ciclo de vida del cliente desde su primera facturación e incorporación a la plataforma hasta su estado de uso activo. El modelo y las políticas diseñadas se aplican exclusivamente a clientes que se encuentran en proceso de *onboarding* o en estado de vigencia, utilizando como base el comportamiento histórico registrado en las bases de datos de la empresa.

En términos de entregables, el trabajo contempla el desarrollo de un prototipo analítico y una carpeta de uso operativa. Esta incluye los scripts de generación del datamart, el cuaderno de *scoring* en Google Colab y los archivos de resultados en Excel. Este material técnico se complementa con la elaboración de tutoriales escritos y la ejecución de sesiones de traspaso de conocimiento dirigidas al equipo de Customer Success y Post-venta, lo que asegura la continuidad de la operación y facilita la futura implementación del modelo.

Respecto a la población analizada, quedan fuera del alcance los clientes que ya se encuentran dados de baja, quienes se utilizan únicamente como historial para el entrenamiento del modelo y mediciones agregadas. Tampoco se consideran en este estudio los potenciales clientes o *leads* que se encuentran en etapas de captación y venta, limitando el análisis a la cartera facturable actual.

En cuanto a la implementación técnica y comercial, se excluye el despliegue productivo en los sistemas centrales de la empresa, la automatización en tiempo real y las integraciones directas con herramientas externas. No se abordan el diseño de campañas de marketing, la compra de medios ni cambios en la estructura de precios o planes comerciales. Los procesos de facturación y cobranza se consideran solo como fuente de información para el análisis, sin intervenir en su operación.

Finalmente, no forma parte del alcance la definición formal de gobierno de datos ni aspectos de cumplimiento regulatorio más allá de lo requerido para el uso interno. Tampoco se desarrollan experimentos causales a gran escala, como pruebas A/B, ni nuevas interfaces de usuario. La capacitación se limita a las sesiones de traspaso técnico con los equipos definidos, sin considerar programas de formación extensiva para otras áreas de la organización.



## 5. Objetivos del proyecto

### 5.1. Objetivo general

Desarrollar un modelo analítico de segmentación conductual para la cartera de clientes de Reservo, que permita diferenciar y gestionar de manera proactiva las estrategias de retención para cada tipo de cliente.

### 5.2. Objetivos específicos

- Obtener un conjunto validado de variables de comportamiento que caractericen a los clientes y permitan predecir el riesgo de abandono y las oportunidades de expansión.
- Construir un repositorio estructurado de datos (datamart) que consolide la información de clientes y habilite el análisis de comportamiento y el desarrollo de modelos predictivos.
- Definir segmentos de clientes definidos y caracterizados en función de sus patrones de comportamiento dinámicos y su evolución en el tiempo.
- Diseñar un portafolio de políticas de negocio personalizadas para cada segmento identificado, orientadas a guiar las acciones del equipo de Customer Success.
- Implementar un prototipo operativo del modelo analítico y del portafolio de políticas en la empresa, con criterios de éxito y lineamientos de adopción.

### 5.3. Indicadores de éxito

- Para el conjunto de variables de comportamiento, la calidad se medirá por el AUC del modelo en datos de prueba; se exigirá un valor igual o superior a 0,70 (Hosmer et al., 2013).
- Para el datamart de clientes, el indicador será la cobertura de la cartera activa: en cada corte mensual al menos un 90 % de los clientes activos debe contar con las variables necesarias para aplicar el modelo (Kumar and Reinartz, 2018).
- Para los segmentos definidos según comportamiento, se evaluará la capacidad de priorización: se espera que el 20 % de clientes con mayor probabilidad concentre al menos un 40 % de las bajas observadas (Inc., 2021).
- Para el portafolio de políticas comerciales, el indicador será disponer de una matriz que relacione segmentos con acciones concretas para Customer Success y Post-venta, que cubra al menos un 80 % de la cartera activa y esté validada con los equipos involucrados (Kumar and Reinartz, 2018).
- Para el prototipo operativo, el indicador será que al menos una persona de Customer Success pueda ejecutar el flujo completo (generar el datamart, correr el notebook de *scoring* y actualizar el Excel de resultados) siguiendo los tutoriales entregados.

## 6. Marco Conceptual

### 6.1. Fundamentos de CRM orientado a valor

El marco conceptual se apoya en la gestión de relaciones con clientes, entendida como un enfoque que combina estrategia, procesos y analítica de datos. La empresa organiza su información comercial y de uso del servicio para tomar decisiones que priorizan la retención y la expansión según el valor económico de cada cliente y su comportamiento observado en el tiempo. El foco se desplaza desde transacciones aisladas hacia una visión de ciclo de vida, donde importa cuánto aporta cada cliente durante toda su relación con la empresa y cómo es posible influir en esa trayectoria (Kumar and Reinartz, 2018).

En un CRM orientado a valor los clientes no se tratan como un conjunto homogéneo. Se agrupan según frecuencia de uso, nivel de pago, respuesta a acciones comerciales y riesgo de baja. Estos grupos permiten diseñar estrategias distintas: algunos segmentos se priorizan por su aporte actual, otros por su potencial de crecimiento y otros porque presentan alta probabilidad de pérdida. La combinación entre valor económico y probabilidad de respuesta se convierte en criterio central para decidir dónde concentrar los recursos de postventa y atención.

Este enfoque es especialmente relevante en modelos de suscripción, donde variaciones pequeñas en la tasa de baja o en la tasa de expansión generan cambios importantes en los ingresos futuros. La literatura de CRM analítico indica que suele ser más eficiente invertir en retener y desarrollar clientes existentes que en adquirir nuevos clientes de manera indiscriminada, siempre que se cuente con métricas claras de valor y con herramientas que permitan identificar a los clientes más relevantes (Gupta et al., 2004, Kumar and Reinartz, 2018). En este contexto, la caracterización conductual de la cartera y su segmentación en función de riesgo y valor se conectan de forma directa con el problema de postventa, ya que permiten avanzar desde un trato uniforme hacia un tratamiento diferenciado y basado en datos.

### 6.2. Herramientas de ingeniería y analítica aplicadas

Las herramientas utilizadas en el proyecto pertenecen al marco de la ingeniería de datos y de la analítica de clientes. En primer lugar se encuentra el concepto de datamart, que proviene de la disciplina de inteligencia de negocios y del diseño de almacenes de datos. Un datamart corresponde a una vista temática de la información de la empresa, organizada en torno a un dominio específico, en este caso la relación con los clientes. Su propósito es ofrecer una base consistente y estable sobre la cual se puedan formular indicadores, análisis y modelos sin depender de los detalles técnicos de los sistemas operacionales (Kumar and Reinartz, 2018).

Sobre esta base se sitúan las herramientas de análisis descriptivo y de exploración de datos, que resumen el comportamiento de la cartera mediante indicadores, curvas de evolución y comparaciones entre grupos. Su aporte es transformar registros dispersos en una representación comprensible de cómo evolucionan los clientes, qué patrones de uso predominan y dónde aparecen diferencias relevantes entre segmentos (Gupta et al., 2004, Kumar and Reinartz, 2018).

En un nivel posterior se ubican los modelos de propensión, que pertenecen al campo de la estadística aplicada y de la minería de datos. Su objetivo es relacionar un conjunto de características de los clientes con la probabilidad de que ocurra un evento de interés, como la baja o la respuesta a una oferta. La regresión logística es uno de los modelos canónicos en este ámbito, ya que permite estimar probabilidades a partir de variables explicativas y ofrece interpretaciones claras sobre el efecto de cada factor (Hosmer et al., 2013). En la gestión de clientes estos modelos se usan para generar scores que ordenan la cartera según nivel de riesgo o probabilidad de respuesta.

La segmentación conductual combina la tradición del marketing analítico con técnicas de clasificación y agrupamiento. En lugar de tratar a todos los clientes por igual, se construyen segmentos que reflejan diferencias en patrones de uso, nivel de ingreso generado y riesgo de baja. Este enfoque se apoya en marcos como las matrices valor versus riesgo, que permiten identificar grupos como clientes de alto valor y alto riesgo o clientes de bajo valor y bajo riesgo. La justificación teórica es que estos segmentos responden de forma distinta a las acciones comerciales y, por lo tanto, requieren tratamientos diferenciados (Gupta et al., 2004, Inc., 2021, Kumar and Reinartz, 2018).

Las reglas de decisión y las políticas comerciales derivadas de la analítica pertenecen al ámbito de la ingeniería de decisiones. Su función es traducir scores, segmentos e indicadores en acciones concretas para los equipos de contacto con clientes. Desde el marco de CRM orientado a valor, estas reglas actúan como puente entre el análisis cuantitativo y la gestión diaria, al establecer qué tipo de intervención corresponde a cada grupo de clientes y bajo qué condiciones se recomienda su aplicación (Kumar and Reinartz, 2018).

### 6.3. Evaluación económica y métricas de desempeño

La evaluación económica del proyecto se apoya en métricas que relacionan el tamaño de la base de clientes, los ingresos recurrentes y el valor económico de cada cuenta. Se define el número de clientes activos como  $N_{\text{activos}}$ , el ingreso mensual recurrente como  $\text{MRR}_{\text{base}}$  y el ingreso promedio por cuenta como  $\text{ARPA} = \text{MRR}_{\text{base}}/N_{\text{activos}}$ . La tasa de baja mensual  $\text{Churn}_{\text{base}}$  es la proporción de clientes activos que dejan de serlo entre dos meses consecutivos y la tasa de expansión mide el porcentaje de clientes que aumentan su pago mensual, con una tasa de referencia  $\text{TasaExp}_{\text{base}}$  y una tasa posterior a la aplicación de políticas  $\text{TasaExp}_{\text{post}}$ .

El valor económico de un cliente se resume mediante el valor de vida del cliente o CLV (Customer Lifetime Value). Bajo supuestos de tasa de baja mensual constante y margen bruto estable, se utiliza la aproximación

$$\text{CLV}_{\text{simple}} = \text{ARPA} \times \text{Margen} \times \frac{1}{\text{Churn}_{\text{base}}},$$

donde Margen es el margen bruto promedio sobre los ingresos recurrentes. La pérdida económica ligada a las bajas se expresa a nivel agregado como ingreso mensual perdido por churn,

$$\text{MRR}_{\text{perdido}} = \text{MRR}_{\text{base}} \times \text{Churn}_{\text{base}}.$$

El beneficio incremental del proyecto se mide comparando la situación antes y después

del uso del modelo. La reducción de la tasa de baja desde  $\text{Churn}_{\text{base}}$  a  $\text{Churn}_{\text{post}}$  genera ingreso retenido adicional  $\Delta\text{MRR}_{\text{retenido}} = \text{MRR}_{\text{base}}(\text{Churn}_{\text{base}} - \text{Churn}_{\text{post}})$  y el aumento de la tasa de expansión desde  $\text{TasaExp}_{\text{base}}$  a  $\text{TasaExp}_{\text{post}}$  genera ingreso adicional por expansión  $\Delta\text{MRR}_{\text{exp}} = \text{MRR}_{\text{base}}(\text{TasaExp}_{\text{post}} - \text{TasaExp}_{\text{base}})$ . Se exige que la combinación de estos efectos, ponderada por el margen y por un horizonte anual, supere el costo del proyecto y permita recuperar la inversión en un plazo razonable. Estos criterios se alinean con la literatura de CRM orientado a valor y con el uso de CLV como eje para priorizar decisiones de retención y expansión (Gupta et al., 2004, Kumar and Reinartz, 2018).

#### **6.4. Elección metodológica para el caso de la empresa**

La elección metodológica se apoya en la necesidad de contar con un marco estándar que ordene el ciclo completo de trabajo analítico. En este contexto se adopta el proceso CRISP-DM (Cross Industry Standard Process for Data Mining), utilizado de forma amplia en proyectos de minería de datos y analítica de negocios. Su objetivo es ofrecer una guía estructurada que conecte la comprensión del negocio, la preparación de datos, el modelamiento y el uso de los resultados en la organización (Chapman et al., 2000).

CRISP-DM define seis fases principales, concebidas como etapas iterativas. En comprensión del negocio se define el problema desde la perspectiva de la organización, se clarifican los objetivos y se fijan criterios de éxito. En comprensión de los datos se identifican las fuentes de información relevantes, se revisan sus estructuras y se exploran sus contenidos, detectando problemas de calidad y formulando hipótesis iniciales. En preparación de los datos se seleccionan variables, se integran fuentes distintas, se limpian registros y se dejan los datos en un formato adecuado para el modelamiento. La fase de modelado consiste en elegir técnicas apropiadas, ajustar parámetros y evaluar resultados preliminares, en interacción con la preparación de datos. En evaluación se analizan métricas cuantitativas y se revisa la coherencia de los modelos con los objetivos de negocio. En despliegue se define cómo se incorporan los resultados en la operación mediante reportes, flujos de scoring, generación de listados y planificación de actualizaciones y monitoreo (Chapman et al., 2000, Kumar and Reinartz, 2018).

Como marco alternativo se considera el proceso de descubrimiento de conocimiento en bases de datos, conocido como KDD (Knowledge Discovery in Databases). Este enfoque organiza el trabajo en etapas de selección, preprocesamiento, transformación, minería de datos y evaluación de patrones, con un énfasis fuerte en las operaciones técnicas de preparación y en la aplicación de algoritmos. La decisión de adoptar CRISP-DM por sobre KDD responde a la naturaleza del problema y a las necesidades de gestión de la empresa. En un contexto de CRM orientado a valor, resulta clave contar con un marco que otorgue un espacio explícito a la comprensión del negocio, a la definición de criterios de éxito y al despliegue operativo de los resultados. CRISP-DM ofrece una estructura más detallada en estas fases, lo que facilita la alineación con los objetivos de gestión de clientes, la comunicación con usuarios de negocio y la planificación de ciclos recurrentes de uso del modelo, mientras que KDD se considera como referencia conceptual para las tareas de minería de datos pero no como guía principal del proyecto (Chapman et al., 2000, Kumar and Reinartz, 2018).

## 7. Metodología

La metodología del proyecto se basa en el proceso CRISP-DM, que organiza el trabajo en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Cada fase agrupa tareas que permiten avanzar desde la definición del problema de gestión hasta la entrega de un prototipo operativo de uso interno (Chapman et al., 2000, Kumar and Reinartz, 2018).

### 7.1. Fase 1: Comprensión del negocio

En esta fase se define el problema desde la perspectiva de la empresa y se alinean las expectativas de los actores involucrados.

- Levantar el contexto de Reservo, incluyendo el modelo de suscripción, el rol de Customer Success y Post-venta, y las principales características de la cartera actual.
- Clarificar el problema de gestión asociado a la baja de clientes y a la necesidad de usar mejor la información disponible para prevenirla y aprovechar oportunidades de expansión.
- Formular el objetivo general y los objetivos específicos del proyecto, asegurando su conexión con causas identificadas del problema de postventa y con necesidades concretas de gestión.
- Definir indicadores de éxito para el modelo, el datamart, la segmentación y el prototipo operativo, y acordar con la empresa el formato de los entregables principales.
- Cuantificar una línea base económica de la cartera mediante el número de clientes activos, el ingreso mensual recurrente, el ingreso promedio por cuenta, la tasa de bajas y la pérdida anual de ingresos asociada, que servirá como referencia para evaluar el impacto del proyecto.

### 7.2. Fase 2: Comprensión de los datos

En esta fase se identifican y se analizan las fuentes de datos que sostendrán el modelo y las segmentaciones.

- Elaborar un inventario de tablas relevantes en la base de datos de Reservo, considerando clientes, uso de la plataforma, cambios de estado administrativos, planes y rubros.
- Revisar la estructura y calidad de cada tabla, analizando llaves, granularidad, cobertura temporal, presencia de datos faltantes, inconsistencias y valores atípicos.
- Definir un calendario de cortes mensuales que represente el estado de la cartera en distintos momentos y sirva de base para la construcción del datamart y de la etiqueta de baja.

- Establecer definiciones operativas de cliente activo, cliente en onboarding, cliente bloqueado y cliente dado de baja, y documentar criterios de inclusión y exclusión de registros.

### **7.3. Fase 3: Preparación de los datos**

En esta fase se construye el datamart analítico y se genera la matriz de características que se utilizará en los modelos.

- Diseñar la estructura del datamart mensual con clave compuesta por identificador de cliente y mes de corte, y definir las uniones necesarias entre tablas de clientes, uso, cambios de estado, planes y rubros.
- Calcular variables de contexto, valor, uso y fricción a partir de las tablas de origen, siguiendo reglas simples y reproducibles que faciliten la interpretación por parte del negocio.
- Definir y construir la etiqueta de baja a horizonte mensual, garantizando que las variables explicativas solo utilicen información disponible hasta el mes de corte y que las bajas consideradas sean efectivas.
- Tratar valores faltantes y valores extremos mediante reglas documentadas, y dejar el conjunto de datos en un formato adecuado para su uso en herramientas de modelamiento.

### **7.4. Fase 4: Modelado**

En esta fase se construyen y se ajustan los modelos que estiman el riesgo de baja y que sirven de base para la segmentación conductual.

- Seleccionar un algoritmo principal de clasificación supervisada, en este caso una regresión logística binaria, y justificar su elección por interpretabilidad y capacidad para producir probabilidades.
- Definir una estrategia de partición temporal en conjuntos de entrenamiento, validación y prueba, que respete el orden de los meses y simule el uso futuro del modelo.
- Implementar un flujo de preprocesamiento que incluya imputación, transformación y codificación de variables, preferentemente encapsulado en un pipeline reproducible.
- Entrenar y ajustar la regresión logística con diferentes configuraciones de parámetros, incorporando técnicas de manejo de desbalance y calibración de probabilidades, y dejando documentadas las opciones evaluadas.

## 7.5. Fase 5: Evaluación

En esta fase se analiza el desempeño del modelo y se conecta con los indicadores de éxito definidos en la comprensión del negocio.

- Evaluar el modelo en el conjunto de prueba utilizando métricas de discriminación y de focalización, como AUC y porcentaje de bajas capturadas en los grupos de mayor riesgo estimado.
- Revisar la calibración de las probabilidades y el comportamiento del modelo en segmentos relevantes, como países, rubros o tipos de plan, para asegurar estabilidad y coherencia.
- Comparar los resultados con los umbrales definidos en los indicadores de éxito y determinar si la calidad del modelo es suficiente para priorizar acciones de retención y expansión.
- Documentar conclusiones y recomendaciones sobre el modelo final, incluyendo sugerencias de ajustes futuros y lineamientos para su monitoreo en uso operativo.

## 7.6. Fase 6: Despliegue y políticas de negocio

En esta fase se define cómo se usarán los resultados del modelo en la operación y se diseña el prototipo operativo que articula analítica y gestión comercial.

- Diseñar un flujo mensual de scoring que permita aplicar el modelo sobre la cartera actualizada, a partir del datamart mensual, y generar tablas con probabilidades de baja, rangos de riesgo y variables de contexto.
- Especificar y construir los artefactos operativos necesarios para el uso del modelo por parte de Customer Success, incluyendo scripts de generación del datamart, un notebook de scoring y archivos de Excel para revisión de resultados.
- Definir una propuesta de segmentación conductual que combine riesgo, valor económico y variables de uso, y asociar a cada segmento un conjunto preliminar de políticas comerciales recomendadas.
- Preparar material de apoyo y lineamientos de adopción, como tutoriales y sesiones de explicación con los equipos de Customer Success y Post-venta, junto con criterios básicos para el monitoreo del prototipo.

## 8. Desarrollo y Resultados

### 8.1. Fase 1: Comprensión del negocio

#### 8.1.1. Contexto de Reservo y modelo de suscripción

En la primera tarea se levantó el contexto del negocio y del modelo de suscripción de Reservo. Se revisaron los planes de servicio, los rubros atendidos y la presencia geográfica, y se confirmó que la empresa opera como proveedor de software como servicio para profesionales y centros de salud pequeños y medianos en Chile, Argentina, Colombia y México, con una cartera cercana a 3.500 clientes activos distribuidos en cuatro planes de precio mensual. Se describió el rol de los equipos de Customer Success y Post-venta en soporte, acompañamiento y retención.

Junto a lo anterior también se analizaron las distribuciones demográficas y comerciales de la cartera. La operación muestra una fuerte concentración geográfica en Chile(anexo E.7), que agrupa al 86 % de los clientes, mientras que el resto se reparte marginalmente entre México, Colombia y otros territorios. A nivel de industria, el rubro de Medicina y Salud abarca el 70 % de la base(anexo E.6), posicionándose como el segmento *core* del negocio, seguido de lejos por Centros de Estética (13 %) y Psicología (9 %). Respecto a la facturación, los planes Básico y Reservo concentran la mayor parte de las suscripciones (77 % combinado), relegando al plan Enterprise a una participación minoritaria del 3 %.(anexo E.8)

El comportamiento cruzado entre planes y rubros revela patrones de riesgo distintivos. Los profesionales de Psicología tienden a optar mayoritariamente por el plan Individual (anexos E.9 y E.10), segmento que presenta una proporción de cuentas bloqueadas (38 %) superior a su participación de clientes activos (20 %) (anexo E.12), lo que sugiere una mayor volatilidad o riesgo de impago en este perfil. Por el contrario, los planes de mayor valor están dominados por el sector médico y muestran tasas de bloqueo considerablemente menores(anexo E.11). Esta segmentación indica que la estabilidad de la suscripción y la probabilidad de bloqueo están correlacionadas tanto con la categoría del plan contratado como con la actividad económica del usuario.

#### 8.1.2. Problema de gestión asociado a la baja de clientes

La segunda tarea clarificó el problema de gestión que motiva el proyecto. Se analizó el ciclo del cliente desde la captación hasta la baja, distinguiendo adquisición, venta, onboarding, cliente activo y cliente inactivo, y se describió el proceso actual de baja y sus señales previas. Las métricas internas mostraron una tasa de bajas mensuales cercana al 2,4 por ciento, un crecimiento moderado de expansiones y esfuerzos de postventa centrados en reaccionar cuando el cliente ya está cerca de abandonar. El problema central se formuló como una falta de caracterización y segmentación conductual de la cartera activa, que impide diferenciar el trato según riesgo y potencial económico.

#### 8.1.3. Definición del objetivo general y de los objetivos específicos

En la tercera tarea se tradujo este diagnóstico en objetivos formales. Se definió como objetivo general diseñar y validar un modelo analítico que estime el riesgo de baja a nivel



de cliente y apoye decisiones de retención y expansión. Se derivaron objetivos específicos para identificar variables de comportamiento relevantes, construir un datamart estructurado, definir segmentos según riesgo y valor, diseñar un portafolio de políticas comerciales diferenciadas e implementar un prototipo operativo integrado al trabajo diario de Customer Success. Cada objetivo se vinculó con causas del problema de postventa identificadas antes.

#### 8.1.4. Indicadores de éxito y definición de entregables

La cuarta tarea definió cómo medir el éxito del proyecto y qué entregables acordar con la empresa. Se fijaron indicadores para la calidad del modelo, como su capacidad para distinguir entre clientes que se dan de baja y los que se mantienen, y para la cobertura del datamart sobre la cartera activa. Se establecieron métricas de focalización, como la concentración de bajas en el grupo de mayor riesgo estimado, y criterios de adopción del prototipo, como que el equipo de Customer Success pueda ejecutar el flujo sin apoyo del estudiante. Se acordaron como entregables clave el datamart mensual documentado, el modelo de riesgo ajustado, la propuesta de segmentación conductual, el portafolio de políticas comerciales asociado y una carpeta de uso con tutoriales y materiales de traspaso para los equipos internos.

#### 8.1.5. Línea base económica y evaluación basada en el Valor de Vida (CLV)

Se utiliza el enfoque de *Customer Lifetime Value* (CLV) para medir el impacto económico del proyecto. Este método cuantifica el valor presente de los flujos de caja futuros que se preservan al evitar la fuga de un cliente, permitiendo valorar el activo financiero que representa la cartera.

##### Determinación de Parámetros y CLV Base

Para el cálculo se utilizan los datos históricos de facturación y se asumen estándares financieros de la industria SaaS (Software as a Service) para los márgenes operativos:

- **Ingreso Promedio (ARPA):** \$58.700 CLP.
- **Churn mensual ( $c$ ):** 2,4 % (0,024).
- **Tasa de descuento anual ( $WACC$ ):** 10 % ( $i \approx 0,008$  mensual).
- **Margen Bruto ( $m$ ):** 70 % (0,70). Este valor corresponde al estándar de la industria SaaS B2B, donde los costos directos de venta (servidores y soporte) son bajos respecto al ingreso recurrente.
- **Margen Neto Estimado:** 10 % (0,10). Se utiliza como referencia conservadora para estimar la utilidad final de la empresa.

El valor del activo cliente se obtiene mediante la fórmula de perpetuidad con crecimiento negativo:

$$CLV = \frac{ARPA \times m}{c + i} \approx \$1.284.062 \quad (1)$$

Cada cliente retenido implica la preservación de un activo valorado en aproximadamente \$1,28 millones.

## Escenarios de Retención y Capacidad Operativa

El modelo predictivo opera con ventanas trimestrales. El impacto económico anual depende de la capacidad del equipo de *Customer Success* para gestionar las alertas generadas en cada ciclo y evitar la baja efectiva.

Se proyectan tres escenarios de efectividad definidos por el volumen trimestral de clientes recuperados. Bajo el supuesto de que un 60 % de las fugas obedece a razones gestionables, y al ponderar este factor por el 80 % de captura del modelo, se establece un potencial máximo de retención del 48 %. La proyección de estas cifras para el año 2025 arroja los siguientes resultados:

- **Pesimista:** Capacidad baja de gestión. Se retienen 20 clientes por trimestre (80 clientes anuales).
- **Neutro:** Capacidad media. Se retienen 30 clientes por trimestre (120 clientes anuales).
- **Optimista:** Capacidad alta. Se retienen 40 clientes por trimestre (160 clientes anuales).

## Evaluación de Impacto Económico

La Tabla 1 detalla el Valor Patrimonial Total preservado en cada escenario. Para dimensionar la magnitud de estas cifras, se debe considerar que la facturación anual de la empresa es cercana a los \$2.400 millones.

Tabla 1: Valor Patrimonial Retenido (CLV acumulado) según escenarios de efectividad operativa.

Escenario	Retención Anual	Valor Preservado (CLV)
Pesimista	80 clientes	\$102.724.960
Neutro	120 clientes	\$154.087.440
Optimista	160 clientes	\$205.449.920

Los resultados indican que, en un escenario de efectividad media (Neutro), el modelo permite retener un valor patrimonial de \$154 millones. Dado el volumen de facturación de la compañía, la preservación de este capital mediante la gestión del churn representa una optimización financiera significativa sobre los activos de la empresa.

## 8.2. Fase 2: Comprensión de los datos

### 8.2.1. Inventario de fuentes y tablas principales

La comprensión de los datos comenzó con un inventario de tablas operacionales sobre clientes, uso de la plataforma y contexto comercial. Se identificó como tabla central a `cliente_cliente`, que contiene la ficha maestra de cada cuenta y permite definir el universo de clientes reales, su localización y su aporte económico. La tabla `ReservoInterno_cambioestadocliente` se definió como fuente principal para reconstruir la historia de estados mediante eventos de tipo 'Baja', 'Bloqueo' y 'Desbloqueo'. El uso básico se obtuvo desde `appointment_ticket` y `cliente_servicio`, que permiten vincular cada cita con un `cliente_id`. El

contexto comercial y el uso avanzado se cubrieron con `cliente_planreservo`, `cliente_rubro`, `Fichas_registroacceso` y tablas de extras, formando un conjunto que resume identidad, valor, uso básico, uso avanzado y contexto de los clientes.

### 8.2.2. Estructura y calidad de las tablas

Una vez definido el inventario se revisó la estructura y calidad de las tablas. En `cliente_cliente` se confirmó la unicidad de `id` y el uso de `demo` y `cuenta_lista` para separar cuentas reales de cuentas de prueba, además de documentar casos de `mensualidad` cero o nula que afectan variables de valor económico. En `ReservoInterno_cambioestadocliente` se verificó la coherencia temporal de los eventos y se observaron patrones de múltiples 'Bloqueo' y 'Desbloqueo' y de 'Baja' seguida de 'Desbloqueo', lo que hizo necesario diferenciar bajas efectivas de bajas revertidas. En las tablas de uso se comprobó la consistencia de las llaves entre citas y clientes y se detectaron tanto periodos de alta actividad como clientes con pocas citas. En planes, rubros, extras y registros de fichas se validaron llaves y nombres, se identificaron clientes sin rubro o plan estándar y se observó concentración del uso avanzado en un subconjunto de la cartera.

### 8.2.3. Calendario de cortes mensuales

La fase incluyó la decisión de trabajar con cortes mensuales como eje temporal común. Se definió un rango de meses de análisis y un corte cercano al inicio de cada mes para observar, en forma regular, el estado y el comportamiento reciente de los clientes. En torno a cada corte se acordó usar ventanas de historia hacia atrás, por ejemplo noventa días, para medir uso de la plataforma y eventos de fricción. Este calendario servirá para responder en cada mes cuántos clientes están activos, qué planes tienen, cuánto utilizan el sistema y qué cambios de estado presentan, y será la base para definir en fases posteriores la ocurrencia de bajas en el mes siguiente.

### 8.2.4. Definiciones operativas de estados de cliente

Por último se establecieron definiciones operativas de estados de cliente. El universo de análisis se fijó en los registros de `cliente_cliente` con `demo = 0`, `cuenta_lista = 1` y `fecha_creacion` válida. Un cliente se considera activo en un mes de referencia si no presenta bajas efectivas previas al corte o si toda baja anterior fue revertida por un evento posterior. Se distinguió entre baja registrada (cualquier evento 'Baja') y baja efectiva, que exige ausencia de 'Desbloqueo' en un plazo razonable, para que las tasas de baja reflejen salidas definitivas. Se definieron también los tramos de onboarding asociados a los primeros meses desde la creación, el estado de bloqueado como señal de fricción y criterios de inclusión o exclusión para casos especiales. Estas definiciones fijan el marco conceptual para la construcción del datamart y de las etiquetas de baja en la fase siguiente.

## 8.3. Fase 3: Preparación de los datos

En esta fase se construye el datamart analítico cliente-mes y se generan las variables que alimentan la matriz de características. El trabajo se organiza en cuatro tareas: diseño del

panel mensual, cálculo de variables de contexto, valor, uso y fricción, definición de la etiqueta de baja a horizonte mensual y tratamiento de valores faltantes y extremos. En el anexo E.1 se puede observar un extracto del código SQL usado para la construcción del datamart.

### 8.3.1. Diseño de la estructura del datamart mensual

La preparación se basa en un script SQL que genera un panel con granularidad `id_cliente`  $\times$  `mes_corte`. El rango temporal se controla con parámetros de inicio y fin del periodo de análisis (`@inicio`, `@fin`), ventanas mínimas para historia de estados y uso (`@hist_inicio`, `@inicio_buf`) y una extensión para cubrir la etiqueta a horizonte T+1 (`@fin_mas1`). Con ellos se construye un calendario de cortes mensuales mediante un CTE recursivo.

El conjunto de clientes candidatos se define a partir de `cliente_cliente`, filtrando registros con `demo = 0`, `cuenta_lista = 1` y `fecha_creacion` válida. La historia de estados se extrae desde `ReservoInterno_cambioestadocliente` para esos clientes entre `@hist_inicio` y `@fin_mas1`, con columnas `id_cliente`, `fecha_cambio` y `tipo_cambio`.

Sobre la combinación `mes_corte-id_cliente` se determina si el cliente está activo en cada corte. Se busca la última 'Baja' previa al corte y se revisa si entre esa fecha y el corte existe un 'Desbloqueo'. Sin 'Baja' previa o con baja revertida el cliente se considera activo; con baja vigente se excluye. Esta estructura se reutiliza luego para unir información de uso, fricción y contexto desde `appointment_ticket`, `cliente_servicio`, `cliente_planreservo`, `cliente_rubro` y `cliente_cliente`, produciendo el panel mensual final.

### 8.3.2. Construcción de variables de contexto, valor, uso y fricción

La segunda tarea consiste en definir las variables que describen a cada cliente en cada mes de corte. Estas variables se calculan en el script principal y se detallan en el documento de variables del datamart. El panel resultante contiene once columnas principales: `id_cliente`, `mes_corte`, `pais`, `rubro`, `plan_nombre`, `arpa_proxy`, `tenure_meses`, `citass_30d`, `recencia_uso_dias`, `bloqueos_90d` y `churn_T1`.

Las variables de identificación y calendario son:

- `id_cliente`: identificador único del cliente, tomado desde `cliente_cliente.id`.
- `mes_corte`: primer día del mes de referencia.

Las variables de contexto se derivan de la información comercial:

- `pais`: país del cliente, proveniente de `cliente_cliente.pais`.
- `plan_nombre`: nombre del plan vigente, obtenido mediante la relación `cliente_cliente.plan_reservo_id`  $\rightarrow$  `cliente_planreservo.id`  $\rightarrow$  `cliente_planreservo.nombre`.
- `rubro`: rubro asociado al plan, calculado como `cliente_planreservo.rubro_id`  $\rightarrow$  `cliente_rubro.id`  $\rightarrow$  `cliente_rubro.nombre`.

Las variables de valor económico son:

- **arpa\_proxy**: aproximación del ingreso mensual por cliente. Se toma desde **cliente\_mensualidad** y se reemplazan los valores cero por nulos mediante **NULLIF(mensualidad, 0)** para evitar confundir promociones o ajustes con ingresos reales.
- **tenure\_meses**: antigüedad del cliente al mes de corte, medida en meses, calculada con **TIMESTAMPDIFF(MONTH, fecha\_creacion, mes\_corte)**.

Las variables de uso reciente de la plataforma se obtienen a partir de las tablas **appointment\_ticket** y **cliente\_servicio**. Primero se agrupa el uso diario por cliente, y luego se calculan:

- **citas\_30d**: número de citas en los 30 días previos al corte. Para cada combinación **mes\_corte, id\_cliente** se suman las citas diarias en el intervalo de 30 días anteriores. Si no hay registros de uso en esa ventana, el resultado se define como cero.
- **recencia\_uso\_dias**: días desde la última cita registrada antes del corte. Se calcula como la diferencia entre **mes\_corte** y la máxima fecha de uso en los 90 días anteriores. Si no hubo uso en los 90 días previos, el valor queda nulo en esta etapa.

La fricción reciente se resume a partir de los cambios de estado administrativos:

- **bloqueos\_90d**: número de bloqueos en los 90 días previos al corte. Para cada cliente activo y corte se cuentan los eventos con **tipo\_cambio = 'Bloqueo'** en la ventana de 90 días anteriores; si no existen bloqueos en ese periodo, se asigna cero.

La combinación de estas variables permite describir, para cada cliente y mes, su contexto comercial, su aporte económico aproximado, su nivel de uso reciente y sus eventos de fricción, manteniendo una estructura mensual uniforme.

### 8.3.3. Definición de la etiqueta de baja a horizonte mensual

La tercera tarea consiste en definir la etiqueta binaria que se usará como resultado en el modelamiento. Esta etiqueta, llamada **churn\_T1**, indica si un cliente activo en un corte determinado presenta una baja efectiva en el mes siguiente.

El cálculo se apoya en dos pasos conceptuales. Primero se identifican las bajas registradas en el mes siguiente a cada corte. Para cada cliente se revisan los eventos de tipo **'Baja'** en **ReservoInterno\_cambioestadocliente** cuya fecha cae entre **mes\_corte** y el mismo día del mes siguiente. Si hay varias bajas en el periodo, se toma la primera fecha de baja del mes.

En segundo lugar se distingue entre bajas registradas y bajas efectivas. Para cada baja registrada se verifica si existe un evento de tipo **'Desbloqueo'** para el mismo cliente con fecha posterior a la baja y anterior al término del mes. Si existe un desbloqueo en esa ventana, la baja se considera revertida y se descarta a efectos de la etiqueta. Si no existe, la baja se clasifica como efectiva.

Con esta definición se construye la etiqueta **churn\_T1**. Para cada cliente activo en un corte dado se pregunta si en el mes siguiente existe una baja efectiva según la regla anterior. Si la respuesta es afirmativa, **churn\_T1** toma valor 1; en caso contrario, toma valor 0. Esta

construcción asegura que todas las variables explicativas de un registro usan solo información observada hasta el mes de corte, mientras que el resultado se define con lo que ocurre en el mes siguiente, evitando fuga temporal.

#### 8.3.4. Tratamiento de valores faltantes y valores extremos

El tipo de datos de las variables se puede encontrar en el anexo E.4. Sobre el panel resultante (anexo E.5) se revisan valores faltantes y extremos. **rubro**, **plan\_nombre** y **pais** presentan nulos que se imputan con una categoría explícita de desconocido(anexo E.3). **arpa\_proxy** tiene casos sin mensualidad válida; se imputan con la mediana por plan o la mediana global y se crea un indicador de imputación. En **recencia\_uso\_dias**, los nulos corresponden a ausencia de uso reciente, se imputan a 90 días y se marca esta situación con un indicador. No se observan correlaciones altas en ninguna combinación de variables tal como se observa en el anexo E.18

Para valores extremos (anexo E.14) se aplican recortes en percentiles altos. **arpa\_proxy** se recorta y se transforma con

$$\log(1 + \text{arpa\_proxy}),$$

manteniendo un indicador de recorte. **tenure\_meses** se capa y se agrupa en bandas de antigüedad (anexo E.13), además de considerar una transformación de raíz cuadrada sobre la versión capada. **citas\_30d** (anexo E.16) se recorta y se transforma con

$$\log(1 + \text{citas\_30d}),$$

junto con indicadores para actividad muy baja y para valores recortados.

**recencia\_uso\_dias** (anexo E.17) se capa en 90 días, se imputan nulos a ese valor y se generan bandas operativas y un puntaje continuo

$$\text{fresh} = \frac{1}{1 + \text{recencia\_uso\_dias}}.$$

**bloqueos\_90d** se trata como conteo discreto pequeño (anexo E.15), capando en tres y generando indicadores para al menos un bloqueo y para dos o más. Todas las imputaciones, recortes y transformaciones se definen usando solo información disponible hasta **mes\_corte** y no utilizan la etiqueta **churn\_T1**, de modo que la matriz final queda lista para el modelado sin fuga temporal.

### 8.4. Fase 4: Modelado

En esta fase se construyen los modelos que estiman el riesgo de baja mensual y que sirven de base para la segmentación conductual. El trabajo se organiza en cuatro tareas: selección del algoritmo, definición de la partición temporal y tamaño muestral, implementación del pipeline de preprocesamiento y entrenamiento con calibración de probabilidades. En el anexo E.2 se puede observar un extracto del código utilizado en Colab/Python.

#### 8.4.1. Selección del algoritmo y definición del problema de clasificación

El problema se formula como una clasificación binaria con horizonte mensual: para cada fila cliente-mes se estima la probabilidad de que un cliente activo presente una baja efectiva en el mes siguiente. La variable objetivo es `churn_T1`, igual a 1 cuando ocurre una baja efectiva en el mes siguiente y 0 en caso contrario. Se elige como algoritmo principal una regresión logística binaria con regularización L2 y `class_weight='balanced'`. El modelo admite variables numéricas transformadas y categóricas codificadas, genera probabilidades entre 0 y 1 y permite interpretar coeficientes y direcciones de efecto, lo que favorece su uso en decisiones de gestión. Métodos más complejos, como ensambles de árboles, se reservan para una etapa futura, privilegiando en esta versión la interpretabilidad y la estabilidad.

#### 8.4.2. Partición temporal y tamaño muestral para entrenamiento y calibración

El panel cliente-mes contiene 43 965 observaciones y 1 073 bajas, con prevalencia cercana a 2,44 %. Se define una partición temporal en tres conjuntos: entrenamiento (19 485 observaciones, alrededor de 497 bajas), desarrollo para calibración (7 821 observaciones, unas 186 bajas) y prueba final (16 659 observaciones, unas 390 bajas), de modo que el entrenamiento use meses más antiguos, el desarrollo meses intermedios y la prueba meses recientes. El modelo utiliza ocho fuentes de información principales: `arpa_proxy`, `citas_30d`, `recencia_uso_dias`, `bloqueos_90d`, `tenure_meses`, `plan_nombre`, `pais` y `rubro`, que tras las transformaciones se descomponen en varias columnas numéricas y dummies. El número de grados de libertad efectivos se sitúa entre 35 y 46, y el criterio de eventos por variable entrega más de 10 eventos por parámetro, lo que respalda la estabilidad de la regresión logística en este contexto de baja prevalencia.

#### 8.4.3. Implementación del flujo de preprocesamiento y del pipeline de modelado

Las transformaciones de la fase de preparación se encapsulan en un pipeline que integra preprocesamiento y regresión logística. Para las variables numéricas (`arpa_proxy`, `citas_30d`, `recencia_uso_dias`, `bloqueos_90d`, `tenure_meses`) se aplican imputaciones, caqueo en percentiles altos, indicadores de nulo y de valor capado, transformaciones continuas como logaritmos o raíces y escalado robusto basado en mediana y rango intercuartílico. Las variables categóricas `plan_nombre`, `pais` y `rubro` se agrupan cuando hay categorías poco frecuentes y luego se codifican mediante one hot. Todas las salidas se combinan en una matriz numérica que alimenta la regresión logística. El conjunto de columnas de entrada se mantiene en `_DEFAULT_FEATURES` y se valida antes de cada ajuste o *scoring* para asegurar consistencia entre bases históricas y futuras.

#### 8.4.4. Entrenamiento, calibración y elección del modelo final

El entrenamiento se realiza ajustando el pipeline completo sobre el conjunto de entrenamiento, usando `_DEFAULT_FEATURES` como entrada y `churn_T1` como salida. En una sola instrucción se aprenden parámetros de imputación, caqueo, escalado, codificación y regresión logística regularizada con ponderación de clases. Luego se evalúa el modelo en el

conjunto de prueba sin calibración adicional, utilizando métricas como área bajo la curva ROC, área bajo la curva precisión–recall y concentración de eventos en los deciles de mayor probabilidad, para verificar su capacidad de ordenar clientes por riesgo. A continuación se calibra la probabilidad envolviendo el pipeline en un clasificador calibrado y usando el conjunto de desarrollo. Según el número de eventos se aplica calibración isotónica o sigmoïdal y se evalúan métricas como Brier score, pendiente e intercepto de calibración y concordancia entre probabilidades predichas y tasas observadas. Se exploran variantes de regularización y subconjuntos de variables y se descartan las que no mejoran discriminación o calibración. El modelo final corresponde a la regresión logística regularizada integrada en el pipeline y envuelta en el calibrador entrenado, que entrega para cada cliente activo en un mes de corte una probabilidad mensual de baja utilizada en las fases de evaluación económica, segmentación conductual y diseño de políticas comerciales.

## **8.5. Fase 5: Evaluación**

En esta fase se analiza el desempeño del modelo y su relación con los indicadores de éxito del negocio, revisando discriminación, focalización, calibración, estabilidad por segmentos y suficiencia para uso operativo.

### **8.5.1. Desempeño global en el conjunto de prueba**

En el conjunto de prueba el modelo calibrado alcanza un ROC–AUC cercano a 0,825 (anexo E.27) y una PR–AUC de aproximadamente 0,10 frente a una tasa base de 0,023, más de cuatro veces un ranking aleatorio (anexo E.20). El primer decil de riesgo presenta una tasa de baja cercana a 11 % y concentra cerca del 48 % de los eventos; el 20 % superior captura alrededor del 68 % de las bajas, el 30 % cerca del 80 %, el 40 % en torno a 86–87 % y el 50 % superior acumula aproximadamente el 91 % de los eventos (anexos E.21 y E.29). Para el 20 % superior (punto de corte cercano a 0,043) el recall es cercano a 0,695 y la precisión a 0,081, con una tasa de eventos más de tres veces la base (anexo E.22). Para un grupo más acotado de unos 600 clientes con mayor score, la precisión se aproxima a 0,148 y el recall a 0,228, configurando un segmento pequeño pero muy cargado de bajas para acciones intensivas (anexo E.30).

### **8.5.2. Calibración de probabilidades y estabilidad por segmentos**

La calibración en prueba muestra un Brier score cercano a 0,022, pendiente de calibración alrededor de 0,89, intercepto próximo a -0,30 y probabilidad media predicha (0,0229) muy similar a la observada (0,0234) (anexo E.28). La tabla de confiabilidad por grupos de probabilidad presenta concordancia razonable, con ligera subpredicción en la cola alta pero en márgenes manejables (anexo E.23). Por plan, Básico y Reservo muestran buena alineación, mientras que Individual presenta una tasa real algo mayor que la estimada y Enterprise diferencias pequeñas con pocos casos (anexo E.25). Por país, Chile mantiene una calibración muy ajustada y los mercados con baja muestra exhiben errores mayores por volumen limitado (anexo E.24). Por rubro, Medicina/Salud y Centro de estética se calibran bien, y en Psicología/Psiquiatría la tasa real supera levemente la estimada, con errores en torno a



0,01, aceptables para uso operativo (anexo E.26).

### **8.5.3. Pruebas de robustez por país y rubro**

Las pruebas de robustez con el mismo modelo entrenado y calibrado muestran que, en segmentos de mayor volumen (global, Chile, Medicina/Salud, Chile más Medicina/Salud y Chile-México-Colombia), el ROC-AUC se mantiene en torno a 0,82-0,83 (anexos E.31, E.32 y E.33), la PR-AUC es similar al valor global y el 20 % superior del ranking captura alrededor del 70 % de las bajas. En México, con menor número de observaciones, el ROC-AUC se sitúa cerca de 0,76(anexo E.34), la PR-AUC triplica la tasa base y el top 20 % captura cerca del 70 % de los eventos (anexo E.37) con precisión sobre 11 %. En Colombia el ROC-AUC cae a valores en torno a 0,62 (anexo E.35), el 20 % superior captura solo cerca del 23 % de las bajas (anexo E.36) y la calibración muestra mayor error, lo que se atribuye al bajo volumen de datos. Se concluye que el modelo global es adecuado como referencia y que mejoras específicas para Colombia como para segmentos más pequeños dependen de contar con más información.

### **8.5.4. Conclusiones sobre suficiencia del modelo para uso operativo**

Los resultados de discriminación y focalización superan los umbrales definidos: el ROC-AUC está por encima del mínimo fijado y el 20 % superior del ranking captura muy por sobre el 40 % de las bajas objetivo, acercándose al 70 %. La calibración global es adecuada, con diferencias promedio pequeñas entre probabilidades predichas y observadas, y la estabilidad por plan, país y rubro es buena en los segmentos principales. Con esta evidencia se considera que el modelo es suficiente para apoyar decisiones de retención y expansión, usando el score mensual de riesgo para priorizar clientes en Customer Success, acompañado de monitoreo regular de métricas y revisiones periódicas de calibración, especialmente en segmentos de menor tamaño.

## **8.6. Fase 6: Despliegue y políticas de negocio**

En esta fase se define cómo se usarán los resultados del modelo en la operación y se concreta un prototipo operativo que conecta el trabajo analítico con la gestión comercial. El diseño considera el flujo mensual de scoring, los artefactos entregados a la empresa, la segmentación conductual asociada a políticas preliminares y los lineamientos de adopción para Customer Success.

### **8.6.1. Flujo mensual de scoring sobre la cartera actualizada**

El flujo de uso parte en la base de datos transaccional de la empresa. Cada vez que se desea actualizar el modelo se ejecuta el script `Datamart Principal Memoria_Act.sql` sobre el esquema `reservov2`, utilizando una herramienta de consulta MySQL como DataGrip conectada mediante túnel SSH a la base productiva. El único parámetro que el usuario debe modificar es la fecha de corte `@fin`, que se fija como el primer día del mes siguiente al periodo que se quiere analizar. El script genera un resultado tabular que contiene, para todos los clientes activos al cierre, las columnas necesarias para el modelo en formato cliente-mes.

Ese resultado se exporta a un archivo **CSV**, que se renombra como **database\_act.csv** y se almacena en la carpeta de trabajo indicada para el modelo. Luego se abre el cuaderno **Colab\_para\_Customer\_Success.ipynb** en Google Colab (anexo F.1), donde se cargan dos archivos: **database\_act.csv** (datamart actualizado) y **database1.csv** (base histórica usada en el entrenamiento). Con estos insumos se ejecuta la celda de producción del cuaderno, que aplica el modelo calibrado sobre el corte seleccionado, calcula las probabilidades de baja a uno y tres meses, asigna rangos cualitativos de riesgo, construye el ranking global y genera archivos de salida en formato **CSV**.

El usuario descarga el archivo de salida elegido, lo renombra como **salida\_modelo\_1.csv** y lo guarda en la carpeta **Colab py**. Finalmente se abre el Excel **Clientes en riesgo de fuga - modelo**, se presiona el botón de “Actualizar todo” en la pestaña de datos y el libro actualiza todas sus consultas de Power Query para leer el nuevo **salida\_modelo\_1.csv**. De este modo el flujo de scoring queda encapsulado en una secuencia de pasos clara: actualización del datamart, ejecución del modelo en Colab y actualización del Excel para la revisión de resultados.

### 8.6.2. Artefactos operativos y carpeta de uso para Customer Success

Para facilitar el despliegue se construye la Carpeta de uso Customer Success (anexo F.3), que organiza todos los elementos necesarios en tres subcarpetas principales. La carpeta Archivo **sql** contiene el script **Datamart Principal Memoria\_Act.sql**, que encapsula las consultas a las tablas de clientes, uso, cambios de estado, planes y rubros. Esta pieza permite regenerar el datamart mensual sin necesidad de escribir código adicional; los usuarios solo deben ajustar la fecha de corte y exportar el resultado.

La carpeta **Colab py** almacena el cuaderno **Colab\_para\_Customer\_Success.ipynb**, el archivo histórico **database1.csv** y el archivo de salida **salida\_modelo\_1.csv**. En el cuaderno se concentran las definiciones del pipeline de preprocesamiento, el modelo calibrado y el módulo de scoring operativo descrito en la fase anterior. El diseño separa explícitamente la etapa de entrenamiento, que no se ejecuta en uso diario, de la etapa de inferencia, que es la única que el equipo debe activar. El archivo **database1.csv** se mantiene fijo como referencia histórica y solo se reemplaza **database\_act.csv** en cada ciclo.

La carpeta Excel de resultado contiene el archivo **Clientes en riesgo de fuga - modelo**, configurado con Power Query para leer siempre el **salida\_modelo\_1.csv** ubicado en **Colab py** (anexo F.2). En el editor de consultas se deja fijada la ruta al archivo y se documenta cómo modificarla en caso de cambio de ubicación. El Excel presenta tablas y vistas filtrables por país, plan, rubro, rango de riesgo y ranking, lo que permite a Customer Success priorizar clientes sin interactuar con código ni con la base de datos. De esta forma, la carpeta de uso reúne en un solo lugar el script SQL, el cuaderno de modelado y el Excel de explotación, junto con los documentos de apoyo asociados.

### 8.6.3. Segmentación conductual y políticas operativas

La salida del modelo se utiliza para definir una propuesta de segmentación conductual basada en combinaciones de riesgo de baja y valor económico aproximado. A partir de las columnas de probabilidad mensual y trimestral, y de los rangos cualitativos de riesgo, se

identifican grupos como alto riesgo y alto valor, alto riesgo y valor medio, riesgo medio con potencial de expansión y riesgo bajo pero valor alto. Estas combinaciones se construyen utilizando la probabilidad a un mes como referencia principal para priorización inmediata y la probabilidad a tres meses para identificar casos donde la baja puede ocurrir dentro del trimestre.

El valor económico se aproxima mediante la variable `arpa_proxy`, derivada de la mensualidad del cliente, y se complementa con la antigüedad en meses y con el plan contratado. Sobre esta base se definen umbrales operativos que permiten ubicar a los clientes en bandas de valor sin necesidad de cálculos adicionales en las herramientas de negocio. La segmentación incluye también información de país y rubro, de modo que las políticas puedan adaptarse a contextos comerciales distintos cuando sea necesario.

El uso del modelo permite optimizar la capacidad instalada del equipo de Customer Success, compuesto por cinco integrantes dedicados a la retención. Actualmente, el proceso de revisar la información de un cliente y contactarlo toma un promedio de dos horas. Con la implementación del modelo, el tiempo de revisión manual de las características del cliente disminuye en un 90 %, ya que el analista dispone de los factores de riesgo ya procesados. Debido a esta priorización, el tiempo total destinado al contacto se reduce en un 50 %, permitiendo que el equipo dedique más tiempo al contacto directo y a la generación de ofertas personalizadas. Bajo estos supuestos, se estima una efectividad de retención de 3 de cada 4 clientes (75 %) en casos donde las razones de fuga son gestionables.

Los motivos de baja observados entre enero de 2023 y julio de 2024 permiten refinar esta segmentación por rubro. En Medicina/Salud, un 33 % de las bajas se explica por cierre de local o quiebra y un 18 % por bloqueos de 40 días, mientras que en Centros de estética la combinación es 27 % por cierre y 19 % por bloqueos. En Psicología/Psiquiatría el cierre de local concentra 31 % de las bajas y el costo beneficio 14 %. En Odontología el principal motivo es la competencia con 23 %, seguido por cierre de local con 21 % y bloqueos con 16 %. En Salón de belleza y barbería se observa una mezcla de cierre de local (18 %), bloqueos (18 %) y cambio a la competencia (13 %), además de un 11 % asociado a costo beneficio. Estos porcentajes entregan una probabilidad relativa de cada motivo dentro de cada rubro y ayudan a priorizar hipótesis de acción cuando un cliente aparece en alto riesgo.

Algo similar ocurre al analizar las razones de baja por plan. En los planes Individual y Básico, el cierre de local representa entre 27 % y 32 % de las bajas y los bloqueos de 40 días entre 16 % y 19 %, seguidos por motivos de costo beneficio en torno a 11–14 %. En el plan Reserva, el cierre de local concentra 35 % de las bajas, la competencia 14 % y los bloqueos 13 %. En el plan Enterprise se observa un peso mayor de temas de producto y requisitos sanitarios: 26 % por cierre de local, 19 % por software y 11 % por falta de permisos sanitarios, junto con porcentajes menores de bloqueos, contactabilidad y costo beneficio (7 % cada uno). Estas diferencias sugieren que la misma probabilidad de baja no debe interpretarse igual en todos los planes, ya que los motivos predominantes cambian con el nivel de servicio.

A partir de esta evidencia, las políticas se ajustan según el decil de riesgo y el perfil del cliente. No se interviene en clientes con riesgo bajo o muy bajo, pues presentan probabilidades de fuga mínimas y permiten un mayor margen de acción si su riesgo aumenta en el siguiente trimestre. En el extremo opuesto, los clientes con riesgo muy alto de fuga al mes siguiente

(T+1) generalmente no son objeto de intervención, pues se consideran casos perdidos donde los esfuerzos no aseguran retención; la excepción a esta regla son los clientes con mensualidades altas o aquellos cuya razón de salida sea la competencia. En el rubro de Medicina, se aplican ofertas agresivas incluso en riesgos extremos para enviar señales competitivas al mercado.

El foco de gestión se concentra en los clientes de riesgo alto y medio, donde existe un margen de maniobra de entre uno y tres meses. Para casos de alto valor en rubros competitivos como Odontología, se recomiendan contactos uno a uno para comparar beneficios y reforzar la propuesta frente a alternativas. En rubros con alta incidencia de bloqueos de 40 días, como Medicina/Salud y planes básicos, se propone el seguimiento sistemático de clientes bloqueados y revisión de causas técnicas. En segmentos de riesgo medio con potencial de expansión, las acciones se orientan a aumentar el uso percibido del software mediante demostraciones de funcionalidades y revisión de reportes. Estas políticas actúan como guía inicial para que el equipo de Customer Success refine sus acciones basándose en la respuesta observada de los clientes.

#### **8.6.4. Material de apoyo, pruebas en vivo y adopción del prototipo**

El despliegue del prototipo se acompaña de un conjunto de tutoriales escritos y materiales audiovisuales que documentan el flujo de principio a fin. Se elabora un tutorial de implementación del modelo que describe la estructura de la Carpeta de uso Customer Success y referencia de manera integrada los documentos específicos para conexión a la base de datos, configuración de Colab y uso del Excel. Se redacta un tutorial de uso diario dirigido al equipo de Customer Success, donde se detallan los pasos para ejecutar el script de datamart, exportar `database_act.csv`, cargar los archivos en Colab, generar las salidas, renombrar `salida_modelo_1.csv` y actualizar el Excel. Este documento enfatiza que los nombres de archivo y las rutas deben mantenerse fijos y que no se deben modificar las consultas internas ni el código del cuaderno (anexos F.4, F.5, F.6).

De manera complementaria se preparan tutoriales específicos para cada herramienta: una guía de conexión de DataGrip a la base de datos de producción mediante túnel SSH, con los parámetros de host, puerto, usuario, llave y base por defecto; una guía de configuración de Google Colab y de carga del cuaderno `Colab_para_Customer_Success.ipynb` en Google Drive; y una guía de uso del Excel de resultados que explica cómo fijar el origen de datos en Power Query y cómo operar el archivo solo con el comando “Actualizar todo”. Cada tutorial incluye versiones en video que muestran el flujo en pantalla, de modo que el equipo pueda seguir los pasos sin necesidad de conocimientos técnicos avanzados.

Durante el cierre del proyecto se realizan pruebas en vivo del flujo completo, utilizando bases actualizadas de producción. Estas pruebas permiten verificar que el datamart se genera correctamente con la fecha de corte deseada, que el cuaderno de Colab acepta los archivos sin errores y que el Excel refleja los cambios una vez cargado el nuevo `salida_modelo_1.csv`. Los resultados de estas ejecuciones se discuten con el equipo de Customer Success, que valida la lectura de las columnas clave y la utilidad de los filtros por riesgo, plan y país.

Como resultado, se deja instalada una primera versión del prototipo operativo, lista para ser utilizada en un periodo inicial de evaluación interna. La empresa planifica usar el modelo

en modo piloto durante los meses de diciembre, enero y febrero, con foco en revisar la coherencia de los clientes priorizados, ajustar las políticas propuestas y detectar necesidades adicionales de segmentación. La implementación completa en los sistemas de la empresa queda fuera del alcance temporal del proyecto, pero todos los elementos necesarios para esa etapa se entregan documentados: scripts, cuadernos, archivos de resultado, tutoriales y recomendaciones de monitoreo. De este modo, el trabajo desarrollado en la memoria deja una base sólida para que la organización avance desde el prototipo hacia una integración más profunda en sus procesos de gestión de clientes.

## 9. Discusiones

### 9.1. Decisiones de diseño del modelo

Las decisiones de diseño del modelo se toman en un contexto donde la empresa requiere resultados accionables y comprensibles. Se elige regresión logística como algoritmo principal porque permite leer coeficientes y efectos de cada variable, lo que facilita que equipos no técnicos entiendan por qué un cliente tiene mayor probabilidad de baja y cómo se vincula ese riesgo con su uso de la plataforma, su plan y sus bloqueos, aun con un posible sacrificio de desempeño frente a modelos más complejos. El horizonte mensual T+1 se alinea con la gestión de cartera basada en facturación y decisiones comerciales mensuales, y la probabilidad T+3 derivada de T+1 entrega una señal adicional para acciones que requieren más tiempo sin entrenar un modelo distinto ni aumentar la complejidad operativa. Se adopta un modelo global para toda la región para reducir la complejidad y aprovechar información de todos los países, pero se mantiene abierta la opción de modelos por país o por rubro cuando la base sea mayor y se requiera más precisión en segmentos específicos. Sin embargo, un factor crítico a considerar para la vigencia futura del diseño es el aumento de precios en los planes ajustados al IPC; esto implica que el nivel de mensualidad con el que fue entrenado el modelo puede variar debido a variables externas, lo que obligará a reentrenar el modelo para incorporar los nuevos niveles de precios y asegurar su validez en el año 2026. El prototipo se implementa con un flujo SQL, Colab y Excel que usa herramientas conocidas y evita cambios en infraestructura, aunque genera dependencia de la disciplina operativa y se plantea como trabajo futuro una integración directa con sistemas de la empresa.

### 9.2. Impacto esperado en la operación y rol de Customer Success

El modelo cambia la forma de priorizar la cartera al entregar probabilidades de baja y un ranking que permiten que Customer Success deje de tratar a todos los clientes activos de la misma manera y concentre tiempo en grupos de mayor riesgo, lo que obliga a revisar la asignación de horas, el número de contactos proactivos posibles al mes y las acciones adecuadas para cada segmento de valor. La segmentación conductual combina riesgo, valor económico y contexto comercial con los motivos de baja observados por rubro y plan para adaptar las intervenciones. No obstante, los escenarios de intervención efectiva enfrentan límites externos, principalmente las acciones de la competencia. Las gestiones del equipo pueden verse condicionadas si los competidores aumentan su poder para atraer clientes, alterando las razones de fuga históricas y generando incertidumbre sobre la efectividad futura de las ofertas. En rubros dominados por cierres de local la capacidad de retención es menor, mientras que en rubros donde destacan competencia o costo beneficio un cliente en alto riesgo lleva a revisar propuesta de valor y precios. El ranking de riesgo ofrece un lenguaje compartido para Customer Success, pero la probabilidad de baja no es el único criterio porque cada acción tiene un costo. La evaluación económica entrega una magnitud referencial, pero se requiere una evaluación más precisa en el futuro que mida el efecto real de las acciones, permitiendo ajustar las prácticas cuando los clientes reinciden en tramos de riesgo a pesar de la gestión.

### 9.3. Comportamiento del cliente y límites del modelo

El modelo se construye con variables de uso de la plataforma, bloqueos, antigüedad, plan, país y rubro, que capturan buena parte del comportamiento observable pero no toda la realidad del cliente. Existen límites importantes al no considerar acciones directas de la competencia ni decisiones subjetivas de administración interna, por lo que pueden aparecer bajas repentinas en clientes que el modelo clasifica como de riesgo bajo y se mantiene un componente inevitable de imprevisibilidad. Un supuesto clave de sensibilidad en el análisis es la independencia de las cuentas; el modelo trata a cada cliente como una unidad independiente, pero existen casos donde diversos clientes pertenecen a un mismo dueño, lo que puede generar diferencias en el paso del análisis a la acción real que deben controlarse manualmente. Otro supuesto sensible es la lógica usada para calcular el churn; debido a limitaciones en la base de datos, pueden existir inconsistencias que dejen a clientes muy antiguos o muy nuevos fuera del análisis, lo que debe corregirse en el futuro para aumentar la precisión sobre la base total real. Intentar representar todas las fuentes de variación con el conjunto actual de datos aumentaría el riesgo de sobreajuste, por lo que la memoria opta por un conjunto moderado de variables con sentido de negocio. Entre las fuentes futuras prometedoras aparece la sistematización de visitas presenciales y el estudio de la oferta de la competencia para interpretar mejor patrones de churn observados solo desde datos internos.

### 9.4. Limitaciones de datos y riesgos en el uso del prototipo

El prototipo depende de la calidad de los registros de bajas y bloqueos, que definen la etiqueta de salida; errores en la codificación de estados o inconsistencias en la identificación de clientes activos pueden distorsionar la estimación del riesgo. Un riesgo operativo relevante es la posibilidad de errores en la salida del modelo al incluir clientes que no sean reales en la base de producción por problemas de limpieza de datos, lo que generaría una pérdida de tiempo valiosa para el equipo al intentar gestionar casos inexistentes. Por ello, se requiere una mejora en la depuración de la base de producción antes de automatizar el proceso. El flujo operativo actual exige varios pasos manuales, como ejecutar el script SQL y actualizar el Excel, y cada paso abre una posible fuente de error humano, de modo que se requiere designar responsables y documentar procedimientos. El entorno de Reservo es dinámico; nuevos competidores, cambios de plan, ajustes de precio o modificaciones en la propuesta de valor pueden alterar los patrones de churn y hacer que el modelo pierda vigencia si no se reentrena, por lo que el proyecto se limita a dejar herramientas funcionales y el mantenimiento futuro dependerá de que la empresa incorpore la gestión activa de estos riesgos y la recalibración periódica en su ciclo de trabajo.

## 10. Conclusiones

El desarrollo del proyecto permite dar por cumplido el objetivo general de dotar a Reservo de un sistema analítico y operativo capaz de gestionar la retención de clientes de manera proactiva. En primera instancia, la validación de variables de comportamiento sugiere que la información transaccional interna resulta suficiente para identificar patrones de riesgo de abandono sin necesidad de recurrir a fuentes externas. El conjunto de indicadores seleccionados, que incluye métricas de valor monetario aproximado, antigüedad, frecuencia de uso reciente y fricción operativa por bloqueos, indica una capacidad predictiva robusta. El rendimiento del modelo alcanzó los umbrales de éxito definidos al inicio del estudio, situándose en un nivel de discriminación que respalda estadísticamente la relación entre estos patrones de conducta y la probabilidad de fuga futura.

En cuanto a la infraestructura de datos, se cumplió el objetivo de construir un datamart estructurado que consolida la visión del cliente en una sola tabla mensual. Este repositorio, generado mediante un script único y parametrizable, alcanza una cobertura efectiva superior al 90 % de la cartera activa en los cortes analizados. La integración de dimensiones de facturación, uso de la plataforma y contexto comercial en una vista unificada no solo alimenta el modelo predictivo actual, sino que habilita al equipo de Customer Success para realizar análisis exploratorios y monitorear la salud de la base de clientes de forma autónoma, resolviendo la carencia de información centralizada previa.

Respecto a la caracterización de la cartera, la segmentación conductual derivada del modelo permitió agrupar a los clientes en rangos de riesgo diferenciados. La capacidad del sistema para concentrar cerca del 70 % del abandono real observado dentro del 20 % de los clientes con mayor puntaje de riesgo sugiere la utilidad práctica de la herramienta para focalizar recursos. Además, el cruce de estos segmentos con los motivos de baja históricos indica patrones comerciales distintivos; se observó que en ciertos rubros predominan causas estructurales, mientras que en otros el riesgo parece estar traccionado por la competencia o la sensibilidad al precio. Este hallazgo aporta una capa de inteligencia de negocio, sugiriendo que una misma probabilidad de baja requiere interpretaciones distintas según el perfil del cliente.

El diseño de un portafolio de políticas comerciales se materializó en una propuesta de reglas de acción que responden a esta segmentación. Las políticas definidas combinan el nivel de riesgo detectado con el valor económico y el contexto del cliente, estableciendo estrategias diferenciadas: contactos personalizados para cuentas de alto valor, campañas digitales automatizadas para segmentos masivos y monitoreo preventivo para casos de riesgo medio. Los lineamientos entregados proporcionan una guía para que el equipo de postventa fundamente sus estrategias en evidencia analítica.

La implementación del prototipo operativo marcó el cumplimiento del objetivo de transferencia tecnológica. La entrega de una carpeta de uso completa asegura que el flujo de trabajo sea reproducible por el equipo de Customer Success de manera autónoma. Las pruebas realizadas indican que el equipo posee la capacidad de generar el datamart, aplicar el modelo sobre un corte mensual y actualizar los resultados, trasladando efectivamente el desarrollo desde un ejercicio académico a una capacidad instalada en la empresa.



Para asegurar la continuidad de estos resultados, la validez del modelo se sustenta bajo supuestos específicos. Se asume que el cliente mantiene un comportamiento estable, lo cual se fundamenta en la revisión de datos históricos de fuga y en consideraciones cualitativas de las áreas de Customer Success y postventa. Asimismo, se proyecta que durante el año 2026 no se realizarán cambios estructurales en la base de datos que afecten la arquitectura del datamart construido. Finalmente, se asume estabilidad en el modelo de negocio de Reserva, donde posibles cambios en los precios de los planes podrían generar variaciones conductuales transitorias antes de retornar a la normalidad operativa.

Bajo este contexto, se establece que el modelo debe ser recalibrado obligatoriamente de forma trimestral para evitar la obsolescencia del entrenamiento. Adicionalmente, se requiere una recalibración extraordinaria ante los siguientes eventos: cambios de precios considerables en la totalidad de los planes, un aumento significativo de clientes en mercados internacionales que permita mejorar la precisión del modelo, o modificaciones en el proceso de bloqueo de suscriptores por impago. Este último punto es crítico, pues un cambio en los periodos definidos para la baja administrativa exigiría modificar los límites y la construcción de la variable de bloqueo en el ingreso de datos.

Finalmente, la integración de estos componentes sustenta la validación de la hipótesis económica del proyecto bajo el enfoque de valor de vida del cliente (CLV). Los análisis de impacto financiero sugieren que la focalización de esfuerzos de retención permite preservar un valor patrimonial significativo. En un escenario de efectividad operativa media, los resultados indican que el modelo contribuye a retener activos por un monto cercano a los \$154 millones anuales. Esta cifra indica que la gestión analítica del churn se constituye como una palanca de rentabilidad crítica, justificando plenamente el esfuerzo de implementación y posicionando a la retención como un eje central de la estrategia comercial de Reserva.

## Referencias

- Biblioteca del Congreso Nacional de Chile (2012). *Ley N° 20.584: Regula los derechos y deberes que tienen las personas en relación con acciones vinculadas a su atención en salud*. Biblioteca del Congreso Nacional de Chile. Accedido en agosto de 2025.
- BusinessWire (2025). Agendapro raises \$35 million growth round led by riverwood capital to give spas, salons, and health & wellness business owners access to ai & automation.
- CEPAL (2024). *ECLAC Launched Digital Development Observatory to Contribute to Latin America and the Caribbean's Digital Transformation*. Economic Commission for Latin America and the Caribbean.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). *CRISP-DM 1.0: Step-by-Step Data Mining Guide*. CRISP-DM Consortium.
- Clínicas de Chile A.G. (2025). *Dimensionamiento del Sector Salud en Chile: Edición 2023*. Clínicas de Chile A.G. URL: <https://www.clinicasdechile.cl/wp-content/uploads/2025/01/Dimensionamiento-2023.pdf>; Accedido: 2025-10-06.
- Encuadrado (2025). *Novedades julio 2025: Nuevo diseño del perfil para Centros*. Encuadrado. URL: <https://www.encuadrado.com/novedades/julio-2025>; Accedido: 2025-10-06.
- Gupta, S., Lehmann, D. R., and Stuart, J. A. (2004). *Valuing Customers*, volume 43. Journal of Marketing Research.
- Healthatom (Medilink) (2025). *Healthatom — Mejoramos la calidad y el acceso a la salud*. Healthatom SpA. URL: <https://www.healthatom.com/>; Accedido: 2025-10-06.
- Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression (3rd ed.)*. Wiley.
- IDB Lab and HolonIQ (2024). *Health Innovation & Technology in Latin America & the Caribbean*. Inter-American Development Bank.
- Inc., S. I. (2021). *Cumulative Gains and Lift Charts (SAS Help Center)*. SAS Institute.
- Jain, S. (2023). *What Is a Good Customer Churn Rate?* ChartMogul. Actualizado: 11 de diciembre de 2023; Accedido: 2025-10-06.
- Kumar, V. and Reinartz, W. (2018). *Customer Relationship Management: Concept, Strategy, and Tools*. Springer.
- Lemmens, A. and Croux, C. (2008). *Bagging and Boosting Classification Trees to Predict Churn*, volume 45. Journal of Marketing Research.
- Munar, N. (2023). *Las healthtech chilenas que están transformando el sector de la salud*. Emol — Portal Pyme.
- Paredes, A. (2025). *Start-Up Chile welcomes a new generation!* Start-Up Chile.
- RESERVO (2025). *Quiénes Somos*. RESERVO. Consultado en agosto de 2025.
- Revista Chilena de Salud Pública (Universidad de Chile) (2025). *Telemedicina en Chile:*

*uso, desarrollo y controversias en una red de salud pública del sur de Chile*. Universidad de Chile. Estudio cualitativo.

Superintendencia de Salud de Chile (2025). *Telemedicina y Salud Mental en Chile*. Superintendencia de Salud de Chile. Documento de trabajo; Ley N. 21.541 formaliza directrices para la práctica, seguridad de datos e interoperabilidad.

Superintendencia de Salud (Gobierno de Chile) (2025). *Chile avanza hacia los 1000 prestadores de salud acreditados en 2026: Un compromiso con la calidad y seguridad del paciente*. Superintendencia de Salud. URL: <https://www.superdesalud.gob.cl/noticias/2025/04/chile-avanza-hacia-los-1000-prestadores-de-salud-acreditados-en-2026-un-compromiso-con-la-calidad-y-seguridad-del-paciente/>; Accedido: 2025-10-06.

World Health Organization (2018). *Classification of digital health interventions v1.0*. World Health Organization.

## Anexo A. Contexto y Problema

### ORGANIGRAMA GENERAL

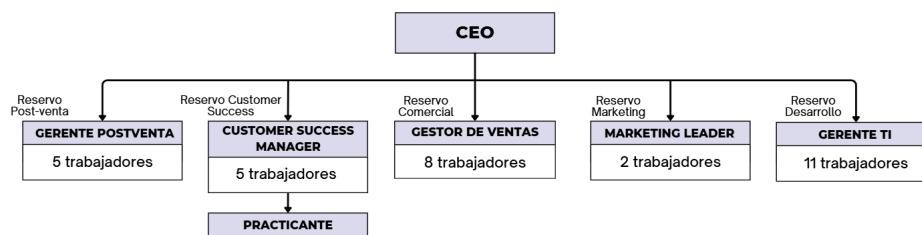


Figura A.1: Organigrama de Reservo, Agosto 2025. Fuente: elaboración propia

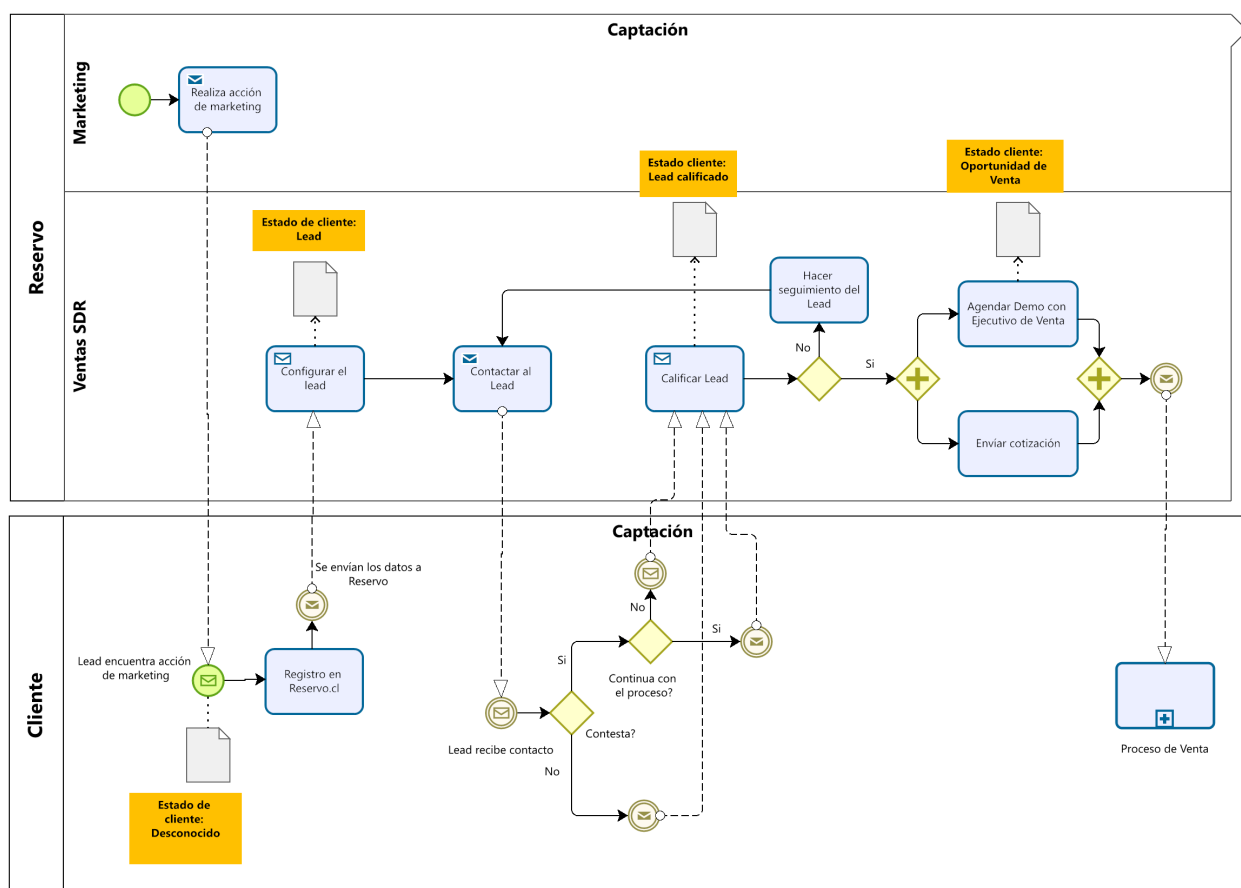


Figura A.2: Etapa de Adquisición en Reservo. Fuente: elaboración propia

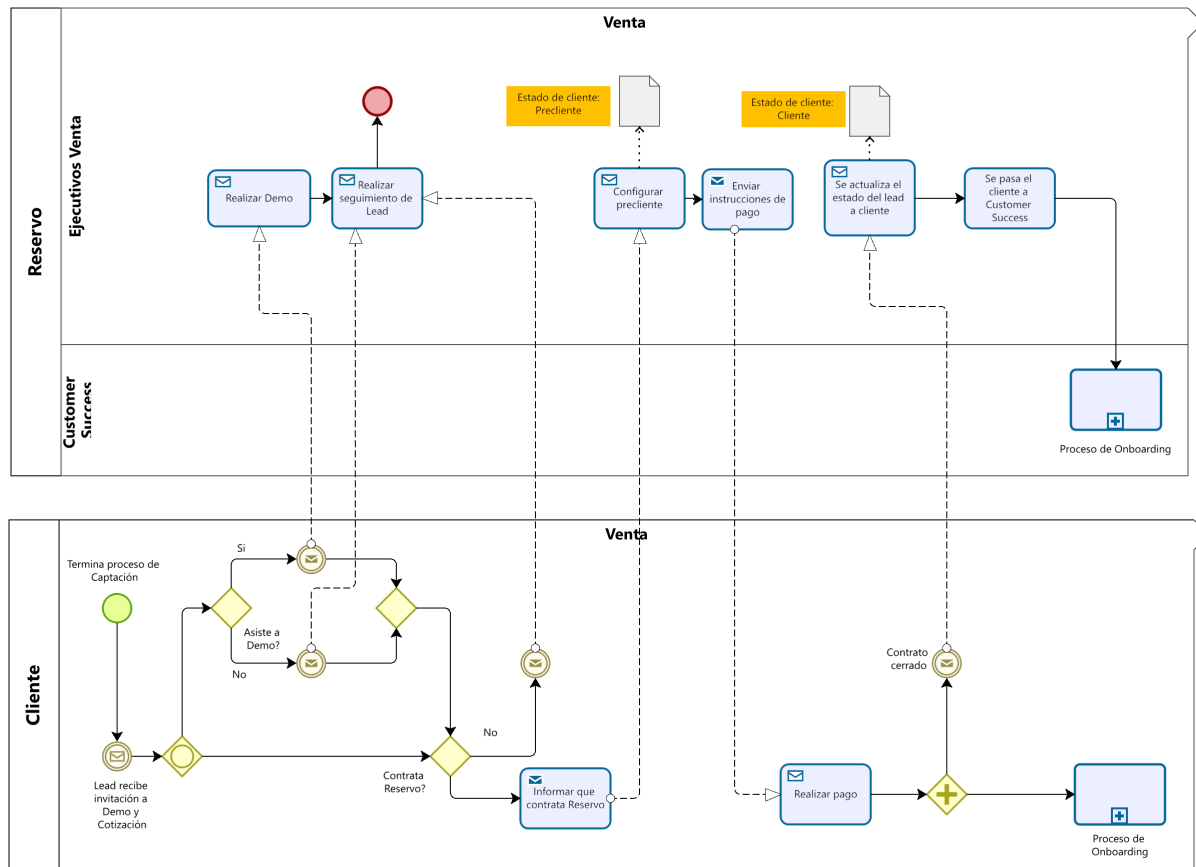


Figura A.3: Etapa de Venta en Reservo. Fuente: elaboración propia

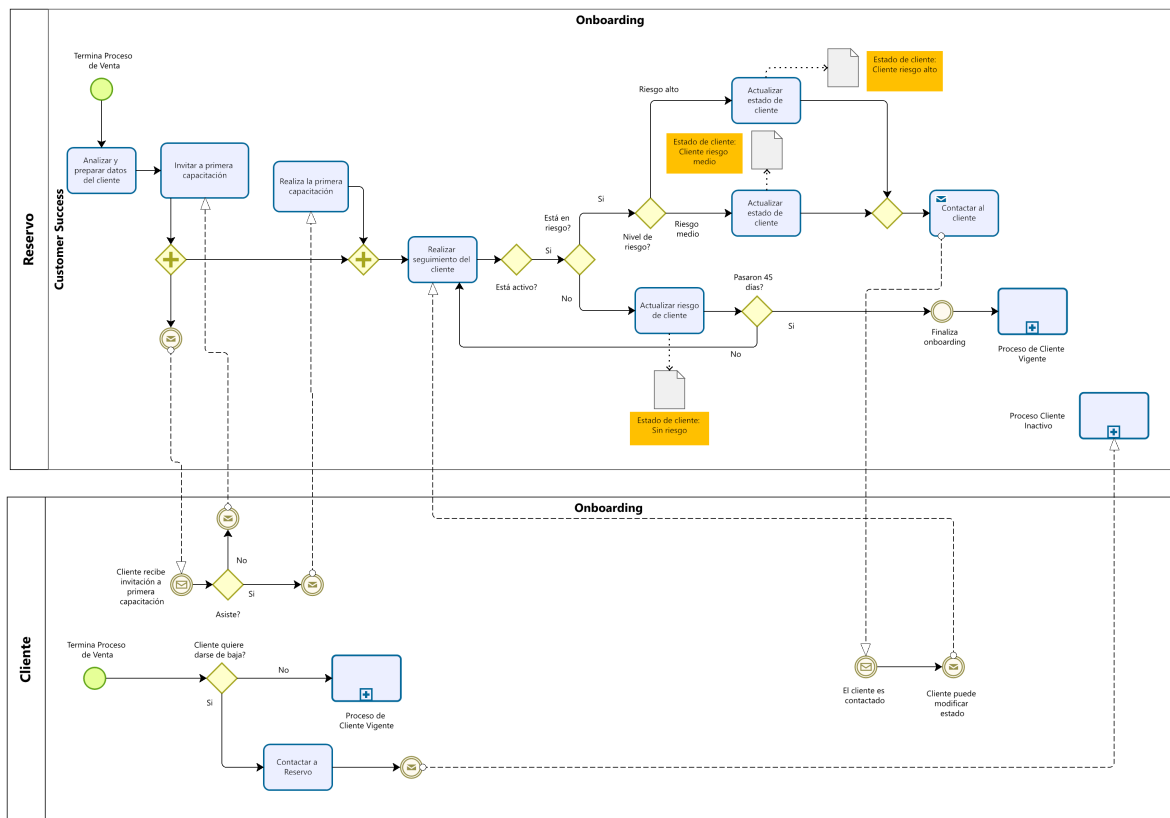


Figura A.4: Etapa de Onboarding del cliente en Reservo. Fuente: elaboración propia

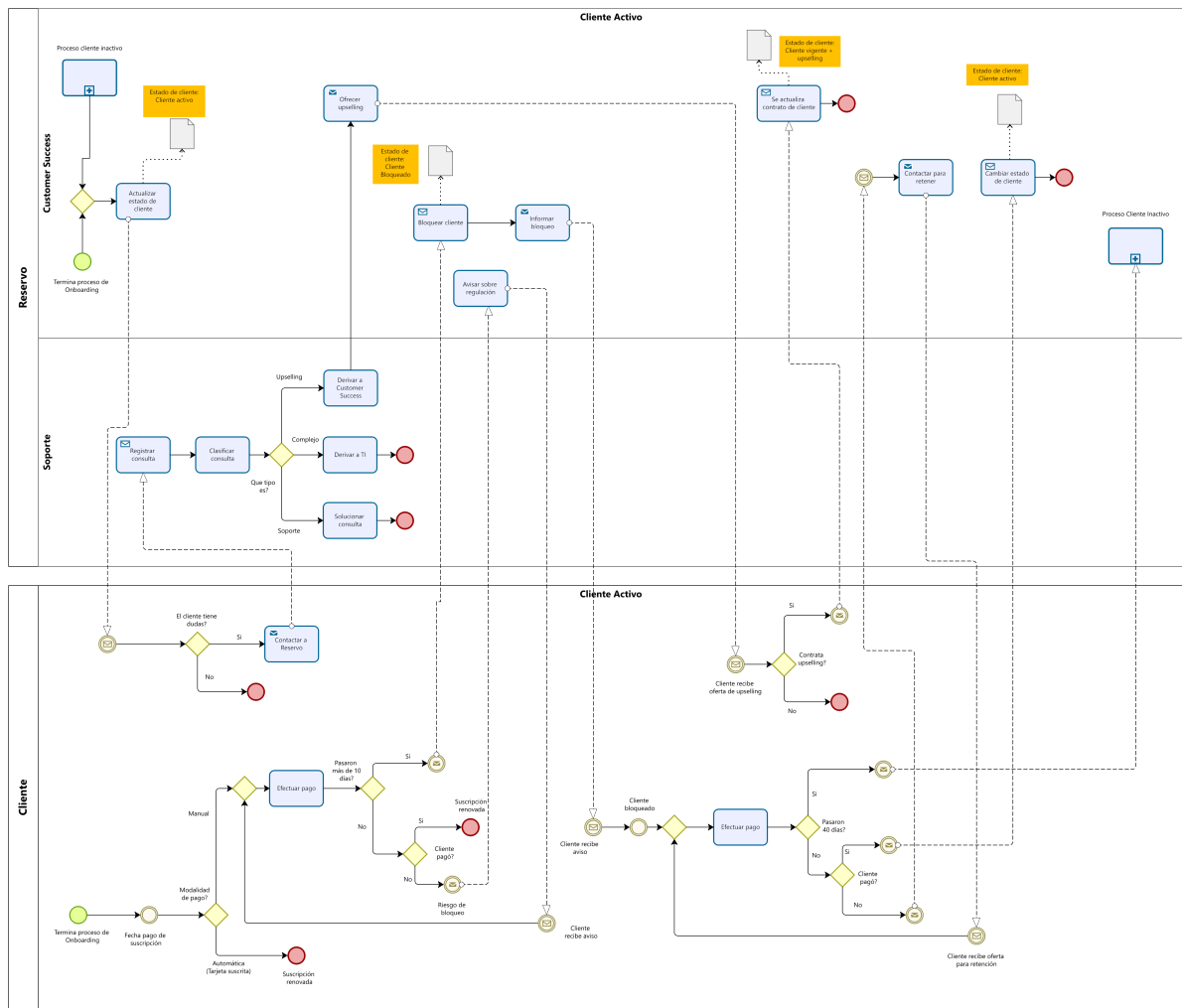


Figura A.5: Etapa de cliente activo en Reservo. Fuente: elaboración propia

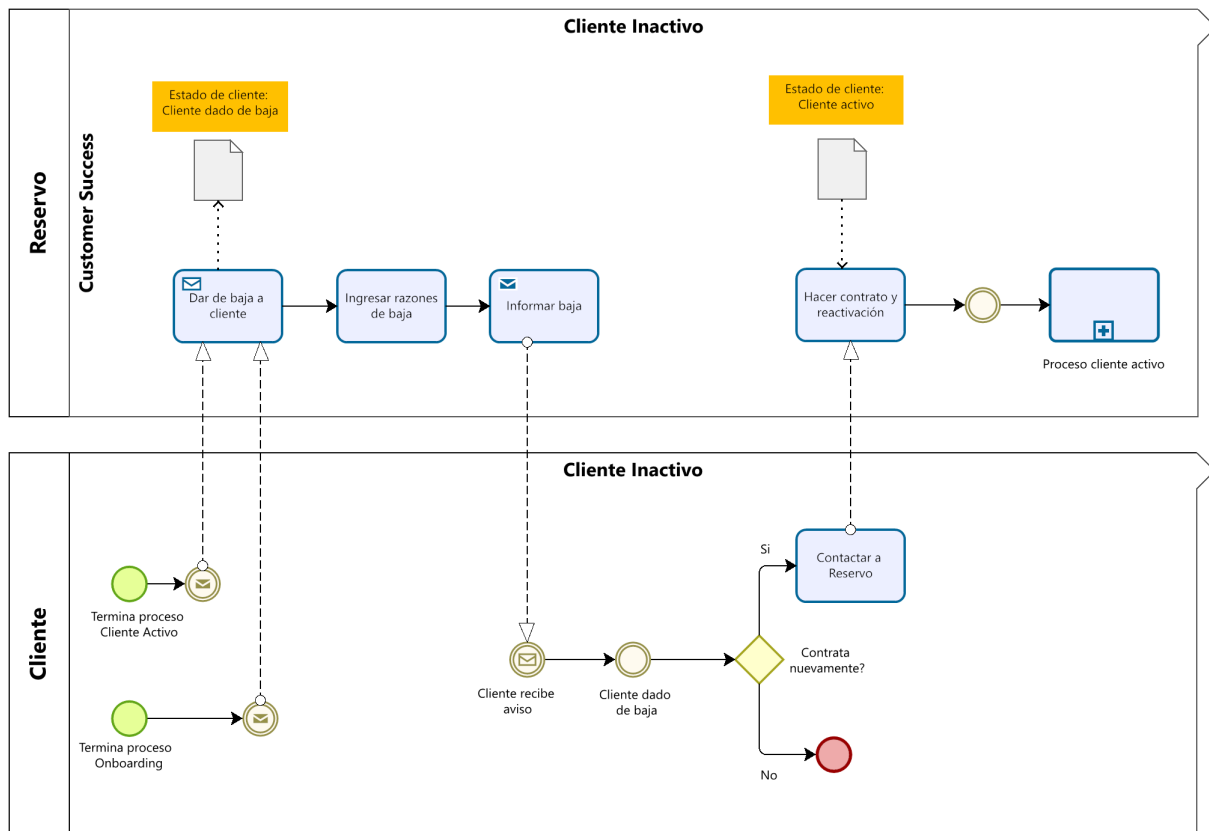


Figura A.6: Etapa de Cliente inactivo en Reservo. Fuente: elaboración propia



## Anexo B. Indicadores

Tabla B.1: Métricas asociadas a clientes. Fuente: Elaboración propia

Estado	Definición	Métricas asociadas	Fórmula / Cálculo
Desconocido	Persona que aún no ha interactuado con la marca.	Visitas al sitio web.	Visitas web: nº de sesiones únicas registradas en la web.
Lead	Persona que mostró interés inicial y dejó sus datos de contacto.	Costo por Lead (CPL).	$CPL = \frac{\text{Gasto en marketing}}{\text{Nº leads generados}}$
Lead calificado	Lead que ha sido validado por el equipo SDR como un cliente potencial.	Tasa de conversión a SQL (Sales Qualified Lead).	$SQL\ rate = \frac{\text{Nº leads calificados}}{\text{Nº leads totales}} \times 100$
Oportunidad de venta	Lead calificado que ha aceptado una demostración o propuesta formal.	Win Rate.	$Win\ rate = \frac{\text{Nº ventas cerradas}}{\text{Nº oportunidades}} \times 100$
Cliente nuevo	Ha contratado el servicio y se ha completado su alta en el sistema.	Costo de Adquisición de Cliente (CAC).	$CAC = \frac{\text{Costo total de Marketing y Ventas}}{\text{Nº clientes nuevos adquiridos}}$
Onboarding (0–90 días)	Cliente nuevo en el proceso de implementación y capacitación inicial.	Tasa de Activación.	$\text{Activación} = \frac{\text{Nº clientes que completan onboarding}}{\text{Nº clientes nuevos}} \times 100$
Cliente activo	Cliente que ha completado el onboarding y usa la plataforma regularmente.	Engagement Score.	Engagement = Frecuencia y profundidad de uso de funciones clave.
Cliente en riesgo	Cliente con bajo uso, quejas recurrentes o problemas de pago.	% de Clientes en Riesgo.	$\% \text{ en riesgo} = \frac{\text{Nº clientes en riesgo}}{\text{Nº clientes activos}} \times 100$
Cliente en expansión	Cliente activo que contrata nuevas funcionalidades, módulos o sedes.	Net Dollar Retention (NDR).	$NDR = \frac{\text{MRR inicial} + \text{Expansión} - \text{Churn}}{\text{MRR inicial}} \times 100$
Cliente Churn	Cliente que ha cancelado su suscripción y no renueva.	Tasa de Churn.	$\text{Churn Rate} = \frac{\text{Nº clientes perdidos}}{\text{Nº clientes al inicio del período}} \times 100$
Valor de Vida del Cliente	Predicción del beneficio neto atribuido a toda la relación futura con un cliente.	Lifetime Value (LTV).	$LTV = (\text{Ticket promedio} \times \text{Recurrencia}) \times \text{Vida del cliente}$

## Anexo C. Tablas de desarrollo

Tabla C.1: Bloqueados y actividad de citas por rubro. Fuente: elaboración propia

Rubro	Número de bloqueados	Prom. citas últimos 3 meses (bloq.)	Prom. citas último año (act.)
Centro de estética	9	16	449
Medicina / Salud	54	20	564
Odontología	2	39	286
Psicología / Psiquiatría	18	6	208
Salón de belleza / Barbería	6	53	544

Tabla C.2: Bloqueados y actividad de citas por plan. Fuente: elaboración propia

Plan	Número de bloqueados	Prom. citas últimos 3 meses (bloq.)	Prom. citas último año (act.)
Individual	36	11	79
Básico	39	16	470
Reservo	16	35	575
Enterprise	3	49	2 524

Tabla C.3: Promedio de mensualidad por rubro (CLP). Fuente: elaboración propia

Rubro	Promedio de mensualidad
Medicina / Salud	\$ 62.566
Psicología / Psiquiatría	\$ 54.731
Centro de estética	\$ 51.366
Salón de belleza / Barbería	\$ 44.906
Odontología	\$ 40.950
<b>Total general</b>	<b>\$ 58.700</b>

## Anexo D. Código

Código D.1: Script SQL para la construcción del Datamart y cálculo de variables.

```

1  /* ===== Parámetros del panel ===== */
2  SET @inicio      := DATE('2024-01-01'); -- primer corte (inicio de mes)
3  SET @fin         := DATE('2025-12-01'); -- último corte (inicio de mes, inclusive)
4  SET @hist_inicio := DATE_SUB(@inicio, INTERVAL 90 DAY); -- ventanas 90d previas
5  SET @inicio_buf  := DATE_SUB(@inicio, INTERVAL 90 DAY); -- buffer de uso
6  SET @fin_mas1    := DATE_ADD(@fin, INTERVAL 2 MONTH); -- cubrir T+1 del ú
    ↪ ltimo
7
8  /* ===== Calendario de cortes ===== */
9  WITH RECURSIVE cortes AS (
10   SELECT @inicio AS corte
11   UNION ALL
12   SELECT DATE_ADD(corte, INTERVAL 1 MONTH)
13   FROM cortes
14   WHERE corte < @fin
15 ),
16
17 /* ===== Clientes candidatos (reales) ===== */
18 candidatos AS (
19   SELECT
20     cc.id AS id_cliente,
21     cc.fecha_creacion AS alta
22   FROM cliente_cliente cc
23   WHERE cc.demo = 0
24     AND cc.cuenta_lista = 1
25     AND cc.fecha_creacion IS NOT NULL
26 ),
27
28 /* ===== Cambios de estado en rango ===== */
29 cambios_src AS (
30   SELECT
31     cec.cliente_id AS id_cliente,
32     cec.fecha AS fecha_cambio,
33     cec.tipo AS tipo_cambio
34   FROM ReservoInterno_cambioestadocliente cec
35   JOIN candidatos ca
36     ON ca.id_cliente = cec.cliente_id
37   WHERE cec.fecha >= @hist_inicio
38     AND cec.fecha < @fin_mas1
39 ),
40

```

```
41 /* ===== BASE cliente x mes por ÚLTIMA ACTIVIDAD PREVIA
    ↳ ===== */
42 base AS (
43     SELECT co.corte, ca.id_cliente
44     FROM cortes co
45     JOIN candidatos ca
46     ON EXISTS (
47         SELECT 1
48         FROM cambios_src cs
49         WHERE cs.id_cliente = ca.id_cliente
50             AND cs.fecha_cambio < co.corte
51     )
52 ),
53
54 /* ===== Activos al corte: última BAJA previa no revertida ===== */
55 activos AS (
56     SELECT b.corte, b.id_cliente
57     FROM base b
58     WHERE NOT EXISTS (
59         SELECT 1
60         FROM cambios_src xb
61         WHERE xb.id_cliente = b.id_cliente
62             AND xb.tipo_cambio = 'Baja'
63             AND xb.fecha_cambio = (
64                 SELECT MAX(x2.fecha_cambio)
65                 FROM cambios_src x2
66                 WHERE x2.id_cliente = b.id_cliente
67                     AND x2.tipo_cambio = 'Baja'
68                     AND x2.fecha_cambio < b.corte
69             )
70         AND NOT EXISTS (
71             SELECT 1
72             FROM cambios_src d
73             WHERE d.id_cliente = b.id_cliente
74                 AND d.tipo_cambio = 'Desbloqueo'
75                 AND d.fecha_cambio >= xb.fecha_cambio
76                 AND d.fecha_cambio < b.corte
77         )
78     )
79 ),
80
81 /* ===== Bajas del mes y churn efectivo ===== */
82 bajas_mes AS (
83     SELECT
84         b.corte AS mes,
85         cs.id_cliente,
86         MIN(cs.fecha_cambio) AS fecha_baja
```

```
87 FROM base b
88 JOIN cambios_src cs
89   ON cs.id_cliente = b.id_cliente
90   AND cs.tipo_cambio = 'Baja'
91   AND cs.fecha_cambio >= b.corte
92   AND cs.fecha_cambio < DATE_ADD(b.corte, INTERVAL 1 MONTH)
93 GROUP BY b.corte, cs.id_cliente
94 ),
95 churn_efectivo AS (
96   SELECT bm.mes, bm.id_cliente
97   FROM bajas_mes bm
98   WHERE NOT EXISTS (
99     SELECT 1
100    FROM cambios_src d
101    WHERE d.id_cliente = bm.id_cliente
102          AND d.tipo_cambio = 'Desbloqueo'
103          AND d.fecha_cambio > bm.fecha_baja
104          AND d.fecha_cambio < DATE_ADD(bm.mes, INTERVAL 1 MONTH)
105   )
106 ),
107
108 /* ===== Etiqueta T+1 para activos ===== */
109 label_churn AS (
110   SELECT
111     a.corte,
112     a.id_cliente,
113     CASE WHEN c.id_cliente IS NOT NULL THEN 1 ELSE 0 END AS churn_T1
114   FROM activos a
115   LEFT JOIN churn_efectivo c
116     ON c.mes = DATE_ADD(a.corte, INTERVAL 1 MONTH) -- T+1 real
117     AND c.id_cliente = a.id_cliente
118 ),
119
120 /* ===== USO: tickets por día (sin duplicados) ===== */
121 uso_dia AS (
122   SELECT
123     cs.cliente_id,
124     DATE(at.fecha) AS fecha_dia,
125     COUNT(DISTINCT at.id) AS citas_dia
126   FROM appointment_ticket at
127   JOIN cliente_servicio cs
128     ON cs.id = at.ofertante_id
129   WHERE at.fecha >= @inicio_buf
130         AND at.fecha < @fin
131   GROUP BY cs.cliente_id, DATE(at.fecha)
132 ),
133
```

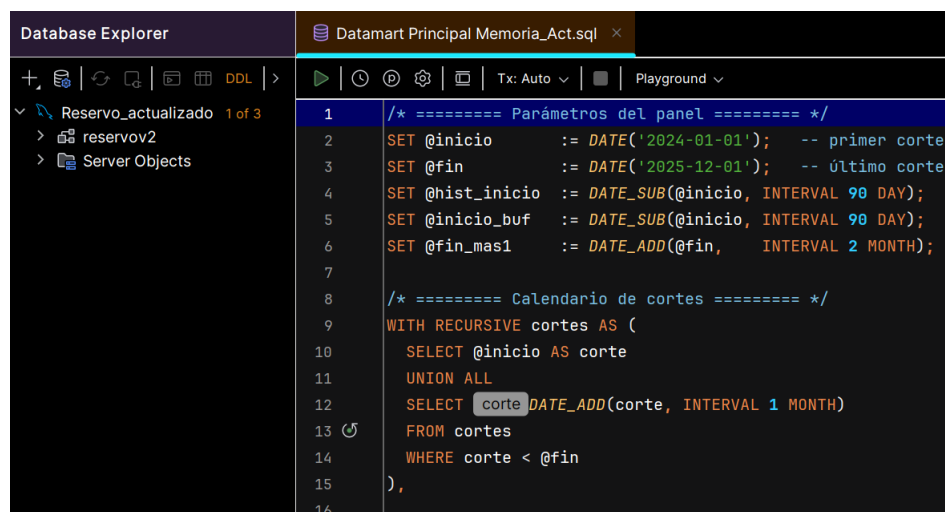
```

134 /* ===== USO LITE por corte: 30d + recencia ===== */
135 uso AS (
136     SELECT
137         a.corte,
138         a.id_cliente,
139         COALESCE(SUM(CASE WHEN ud.fecha_dia >= DATE_SUB(a.corte, INTERVAL 30
140             ↪ DAY)
141                 AND ud.fecha_dia < a.corte THEN ud.citas_dia END),0) AS citas_30d,
142         DATEDIFF(a.corte, MAX(ud.fecha_dia)) AS recencia_uso_dias
143     FROM activos a
144     LEFT JOIN uso_dia ud
145         ON ud.cliente_id = a.id_cliente
146         AND ud.fecha_dia < a.corte
147         AND ud.fecha_dia >= DATE_SUB(a.corte, INTERVAL 90 DAY)
148     GROUP BY a.corte, a.id_cliente
149 ),
150 /* ===== Fricción LITE: solo bloqueos_90d ===== */
151 friccion AS (
152     SELECT
153         a.corte,
154         a.id_cliente,
155         COALESCE(SUM(CASE WHEN cs.tipo_cambio = 'Bloqueo'
156             AND cs.fecha_cambio >= DATE_SUB(a.corte, INTERVAL 90 DAY)
157             AND cs.fecha_cambio < a.corte THEN 1 END),0) AS bloqueos_90d
158     FROM activos a
159     LEFT JOIN cambios_src cs
160         ON cs.id_cliente = a.id_cliente
161         AND cs.fecha_cambio < a.corte
162         AND cs.fecha_cambio >= DATE_SUB(a.corte, INTERVAL 90 DAY)
163     GROUP BY a.corte, a.id_cliente
164 ),
165
166 /* ===== Valor / contexto (añade país y rubro) ===== */
167 valor AS (
168     SELECT
169         a.corte,
170         a.id_cliente,
171         cc.pais,
172         cpr.nombre AS plan_nombre,
173         cr.nombre AS rubro,
174         NULLIF(cc.mensualidad,0) AS arpa_proxy,
175         TIMESTAMPDIFF(MONTH, cc.fecha_creacion, a.corte) AS tenure_meses
176     FROM activos a
177     JOIN cliente_cliente cc
178         ON cc.id = a.id_cliente
179     LEFT JOIN cliente_planreserva cpr

```

```
180     ON cpr.id = cc.plan_reservo_id
181 LEFT JOIN cliente_rubro cr
182     ON cr.id = cpr.rubro_id
183 )
184
185 /* ===== Panel final LITE ===== */
186 SELECT
187     v.id_cliente,
188     v.corte      AS mes_corte,
189     v.pais,
190     v.rubro,
191     v.plan_nombre,
192     v.arpa_proxy,
193     v.tenure_meses,
194     u.citas_30d,
195     u.recencia_uso_dias,
196     f.bloqueos_90d,
197     l.churn_T1
198 FROM valor v
199 LEFT JOIN uso      u ON u.corte = v.corte AND u.id_cliente = v.id_cliente
200 LEFT JOIN friccion  f ON f.corte = v.corte AND f.id_cliente = v.id_cliente
201 LEFT JOIN label_churn l ON l.corte = v.corte AND l.id_cliente = v.id_cliente
202 WHERE v.corte < @fin;
```

## Anexo E. Imágenes de desarrollo

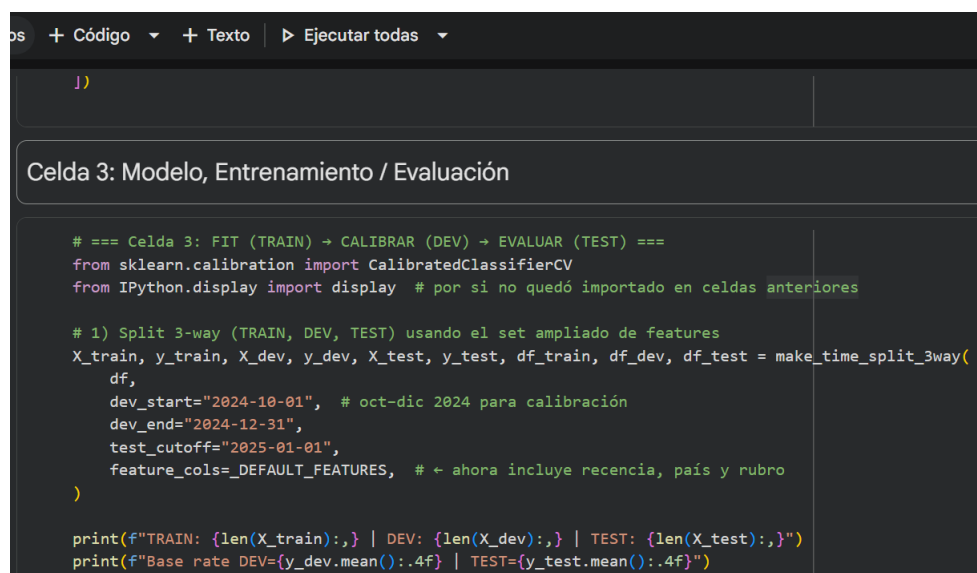


```

1  /* ===== Parámetros del panel ===== */
2  SET @inicio      := DATE('2024-01-01'); -- primer corte
3  SET @fin         := DATE('2025-12-01'); -- último corte
4  SET @hist_inicio := DATE_SUB(@inicio, INTERVAL 90 DAY);
5  SET @inicio_buf  := DATE_SUB(@inicio, INTERVAL 90 DAY);
6  SET @fin_mas1    := DATE_ADD(@fin, INTERVAL 2 MONTH);
7
8  /* ===== Calendario de cortes ===== */
9  WITH RECURSIVE cortes AS (
10     SELECT @inicio AS corte
11     UNION ALL
12     SELECT DATE_ADD(corte, INTERVAL 1 MONTH)
13     FROM cortes
14     WHERE corte < @fin
15 ),
16

```

Figura E.1: Extracto código SQL construcción de datamart. Fuente: elaboración propia



```

# == Celda 3: FIT (TRAIN) → CALIBRAR (DEV) → EVALUAR (TEST) ==
from sklearn.calibration import CalibratedClassifierCV
from IPython.display import display # por si no quedó importado en celdas anteriores

# 1) Split 3-way (TRAIN, DEV, TEST) usando el set ampliado de features
X_train, y_train, X_dev, y_dev, X_test, y_test, df_train, df_dev, df_test = make_time_split_3way(
    df,
    dev_start="2024-10-01", # oct-dic 2024 para calibración
    dev_end="2024-12-31",
    test_cutoff="2025-01-01",
    feature_cols=_DEFAULT_FEATURES, # + ahora incluye recencia, país y rubro
)

print(f"TRAIN: {len(X_train):,} | DEV: {len(X_dev):,} | TEST: {len(X_test):,}")
print(f"Base rate DEV={y_dev.mean():.4f} | TEST={y_test.mean():.4f}")

```

Figura E.2: Extracto de código modelo Python/Colab. Fuente: elaboración propia



Nulos por variable:

	n_nulos	pct_nulos
rubro	2426	0.06
recencia_uso_dias	2178	0.05
arpa_proxy	137	0.00
plan_nombre	73	0.00
pais	0	0.00
id_cliente	0	0.00
mes_corte	0	0.00
tenure_meses	0	0.00
citas_30d	0	0.00
bloqueos_90d	0	0.00
churn_T1	0	0.00

Figura E.3: Porcentaje de nulos por variable. Fuente: elaboración propia

```
Tipos de datos de las variables:
id_cliente          int64
mes_corte           datetime64[ns]
pais                category
rubro               category
plan_nombre         category
arpa_proxy          float64
tenure_meses        int64
citas_30d           int64
recencia_uso_dias   float64
bloqueos_90d        int64
churn_T1            int64
dtype: object
```

Figura E.4: Tipos de datos de las variables. Fuente: elaboración propia

Estadísticas numéricas:										
	count	mean	std	min	25%	50%	75%	90%	95%	max
arpa_proxy	43828.00	57917.24	88203.36	1.00	40000.00	40000.00	60000.00	100000.00	130000.00	3087939.00
tenure_meses	43965.00	27.08	23.44	0.00	8.00	21.00	41.00	62.00	75.00	128.00
citas_30d	43965.00	381.01	685.80	0.00	41.00	173.00	435.00	939.00	1430.80	13791.00
recencia_uso_dias	41787.00	4.19	10.55	1.00	1.00	1.00	2.00	6.00	19.00	90.00
bloqueos_90d	43965.00	0.31	0.73	0.00	0.00	0.00	0.00	1.00	2.00	5.00
churn_T1	43965.00	0.02	0.15	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Figura E.5: Estadísticas descriptivas de variables numéricas. Fuente: elaboración propia

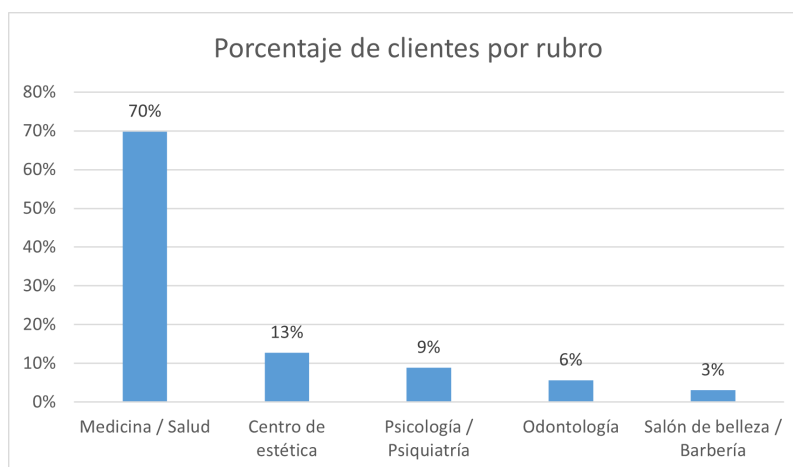


Figura E.6: Porcentaje de clientes por rubro. Fuente: elaboración propia

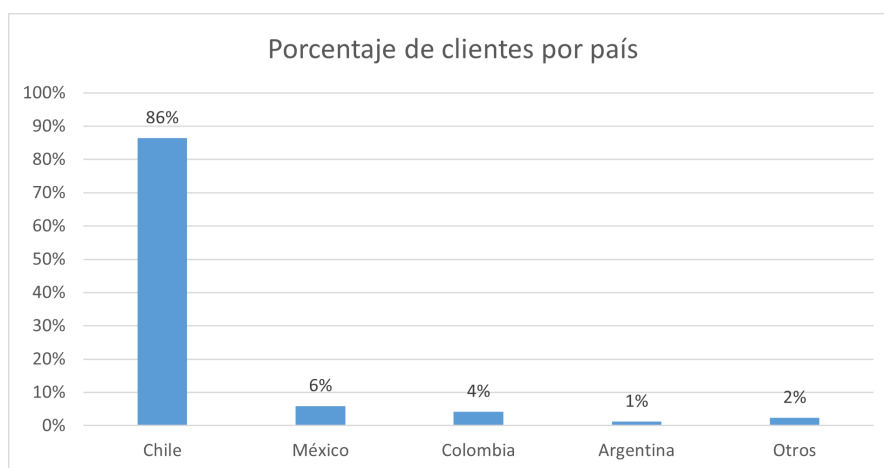


Figura E.7: Porcentaje de clientes por país. Fuente: elaboración propia

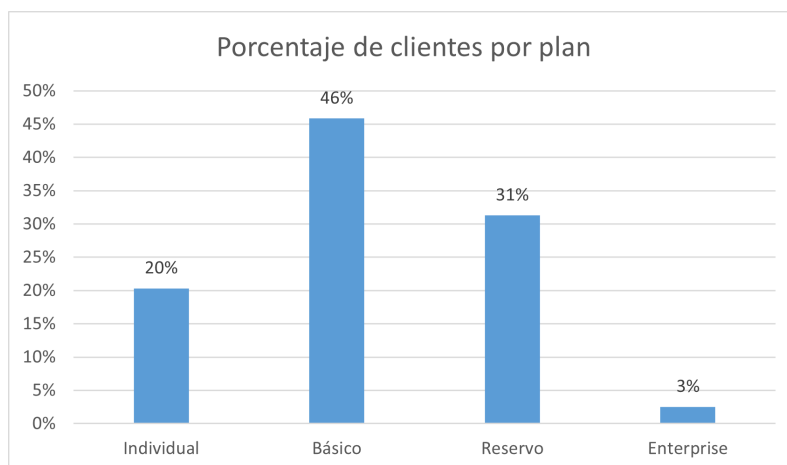


Figura E.8: Porcentaje de clientes por plan. Fuente: elaboración propia

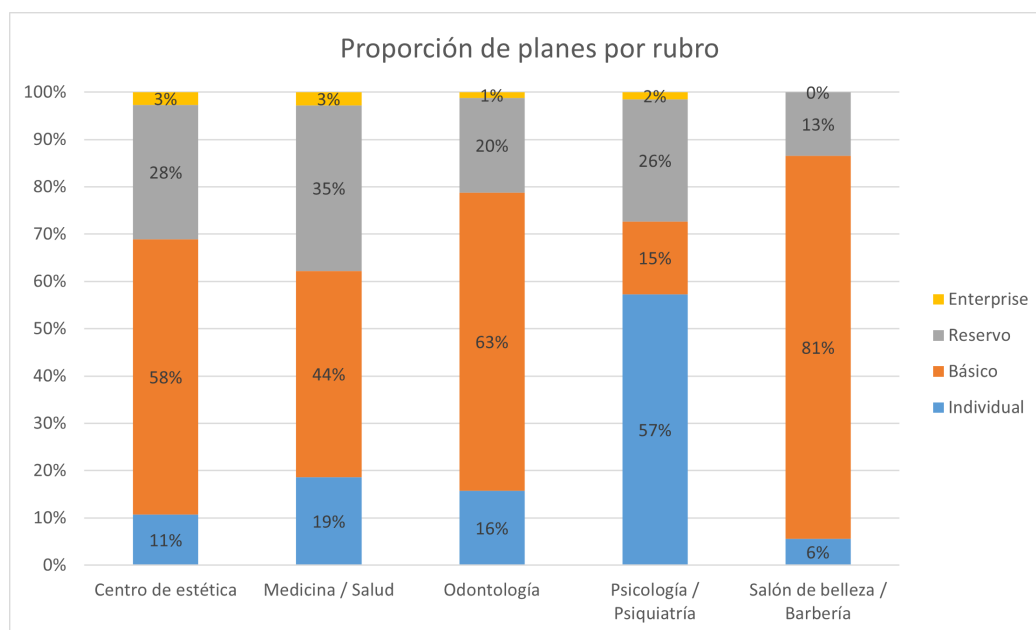


Figura E.9: Proporción de planes por rubro. Fuente: elaboración propia

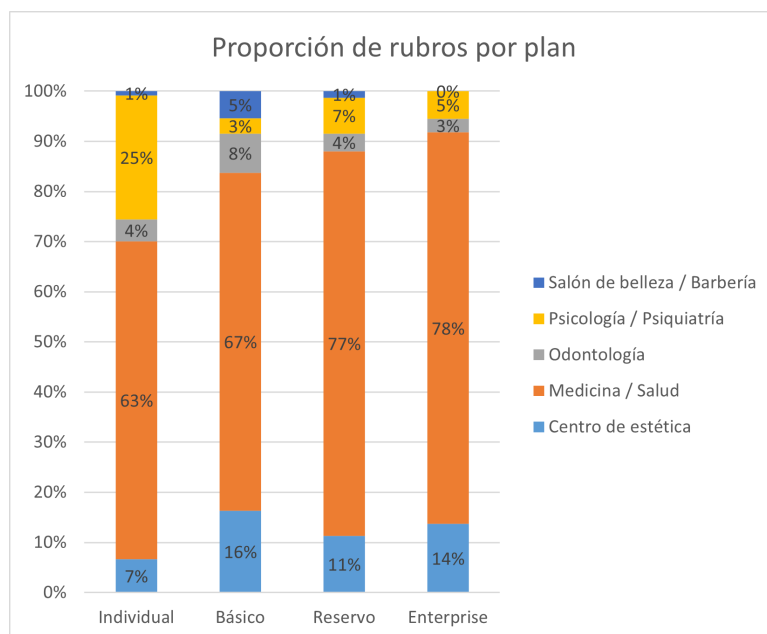


Figura E.10: Proporción de rubros por plan. Fuente: elaboración propia

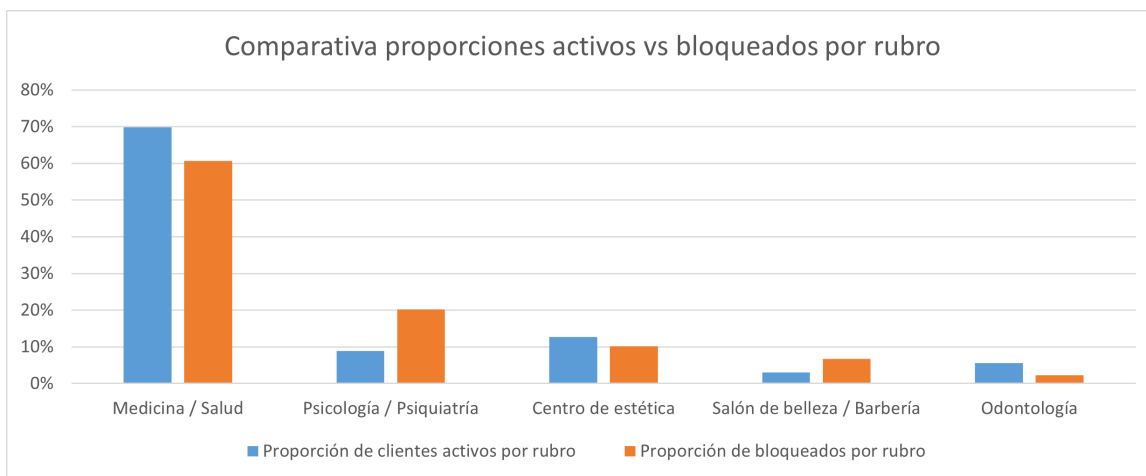


Figura E.11: Comparativa de proporciones de activos vs bloqueados por rubro. Fuente: elaboración propia

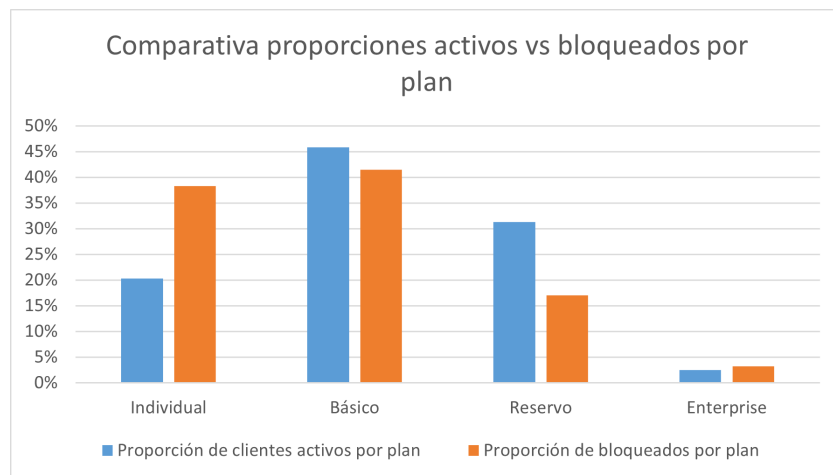


Figura E.12: Comparativa de proporciones de activos vs bloqueados por plan. Fuente: elaboración propia

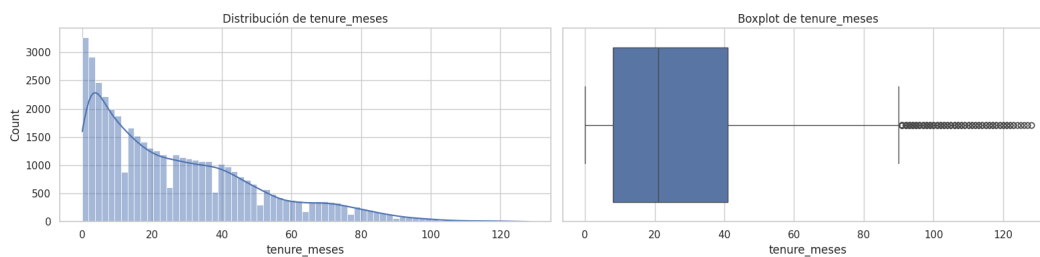


Figura E.13: Distribución y boxplot de *tenure\_meses*. Fuente: elaboración propia

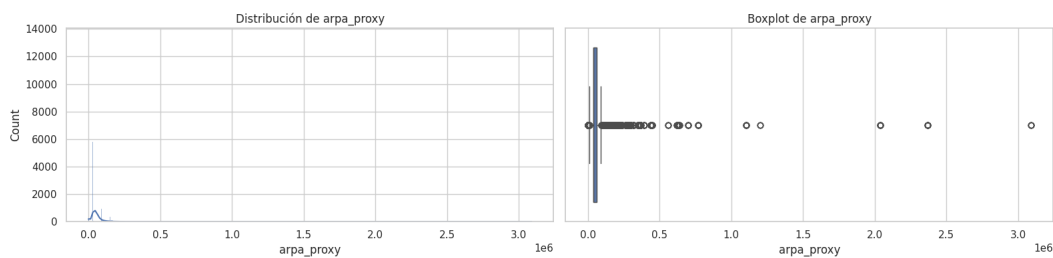


Figura E.14: Distribución y boxplot de *arpa\_proxy*. Fuente: elaboración propia

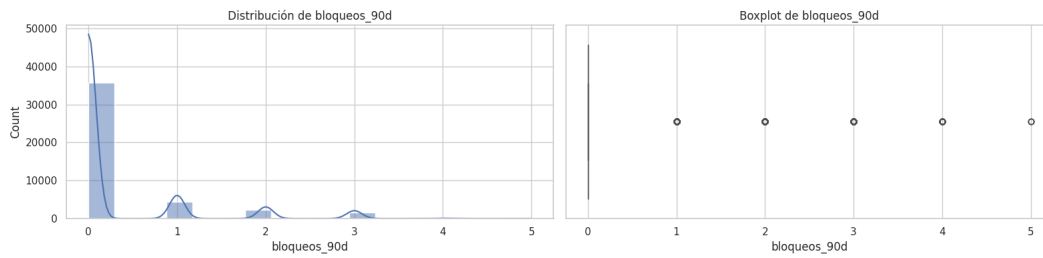


Figura E.15: Distribución y boxplot de *bloqueos\_90d*. Fuente: elaboración propia

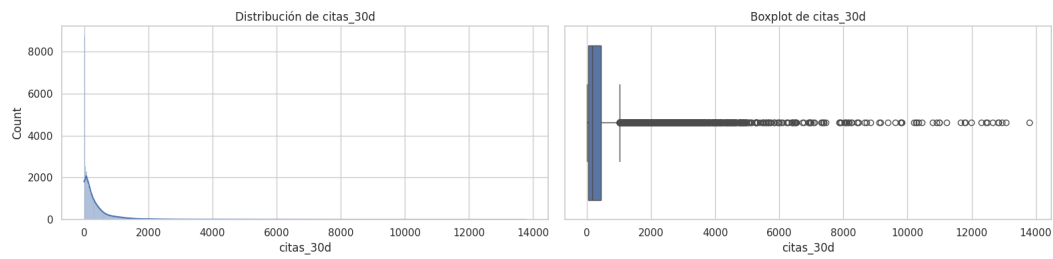


Figura E.16: Distribución y boxplot de *citas\_30d*. Fuente: elaboración propia

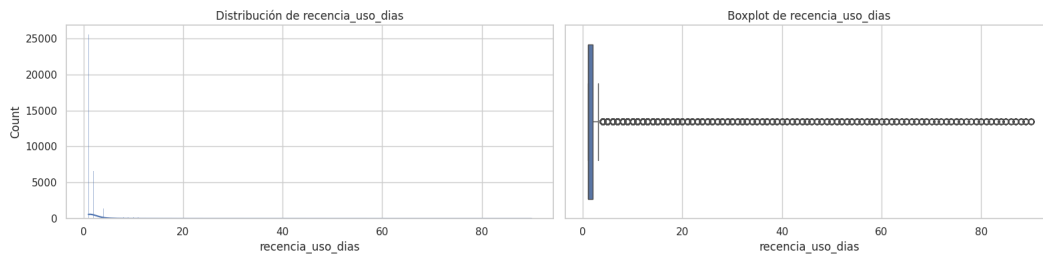


Figura E.17: Distribución y boxplot de *recencia\_uso\_dias*. Fuente: elaboración propia

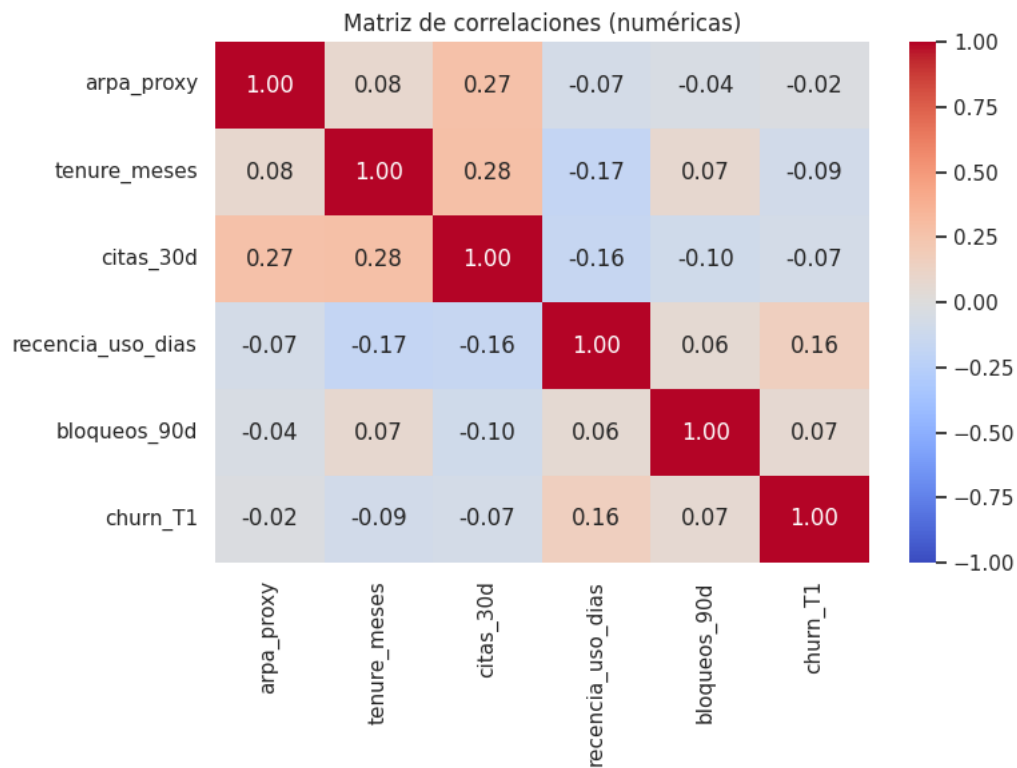


Figura E.18: Matriz de correlaciones entre variables numéricas. Fuente: elaboración propia

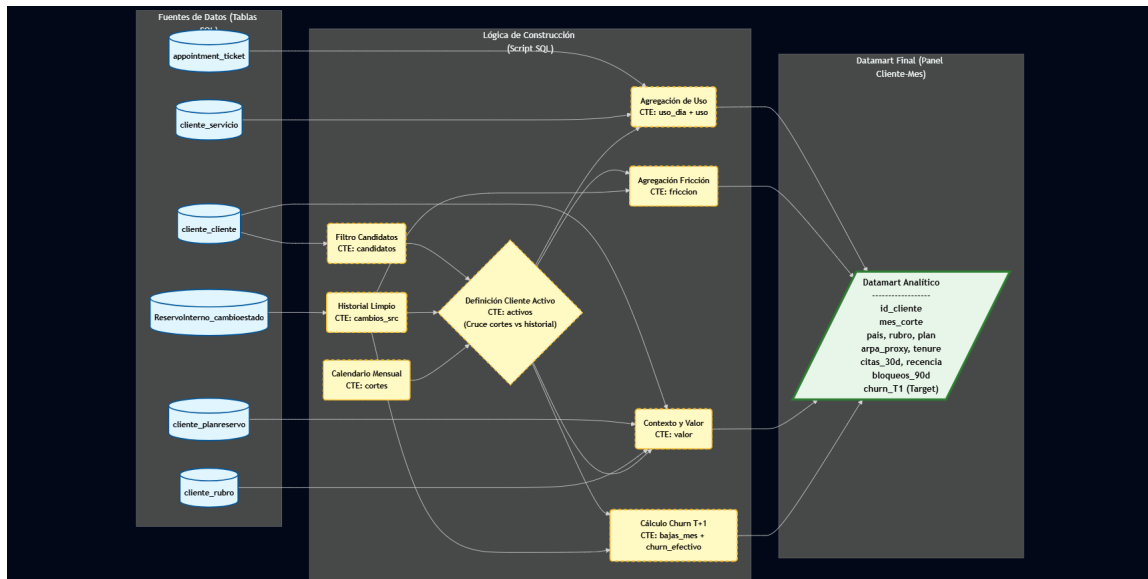


Figura E.19: Diagrama de Datamart construido. Fuente: elaboración propia

```

TRAIN: 19,485 | DEV: 7,821 | TEST: 16,659
Base rate DEV=0.0238 | TEST=0.0234
[Sin calibrar] PR-AUC=0.1122 | ROC-AUC=0.8260
Calibración aplicada con método: isotonic
[Calibrado] PR-AUC=0.0996 | ROC-AUC=0.8246
Tasa base (test): 0.0234

```

Figura E.20: Resultado python capacidad de discriminación del modelo.  
Fuente: elaboración propia

	decile	n	events	avg_score	event_rate	lift	cum_n	cum_events	cum_capture_pct	score_cutoff
0	1	1666	189	0.10	0.11	4.85	1666	189	0.48	0.08
1	2	1666	78	0.06	0.05	2.00	3332	267	0.68	0.04
2	3	1666	45	0.03	0.03	1.15	4998	312	0.80	0.03
3	4	1666	24	0.01	0.01	0.62	6664	336	0.86	0.01
4	5	1666	20	0.01	0.01	0.51	8330	356	0.91	0.00
5	6	1665	5	0.00	0.00	0.13	9995	361	0.93	0.00
6	7	1666	12	0.00	0.01	0.31	11661	373	0.96	0.00
7	8	1666	9	0.00	0.01	0.23	13327	382	0.98	0.00
8	9	1666	5	0.00	0.00	0.13	14993	387	0.99	0.00
9	10	1666	3	0.00	0.00	0.08	16659	390	1.00	0.00

Figura E.21: Resultado python captura acumulada. Fuente: elaboración propia

```

Top 20% -> K=3332 | Recall=0.695 | Precision=0.081 | Cutoff=0.042705
Top N -> K=600 | Recall=0.228 | Precision=0.148 | Cutoff=0.100000
Umbral sugerido (top 20%): 0.042705

```

Figura E.22: Resultado python métricas operativas. Fuente: elaboración propia



Reliability table con 10 bins:

	bin	n	mean_pred	event_rate	abs_err	rel_err
0	(-0.001, 0.00105]	3005	0.00	0.00	0.00	0.69
1	(0.00105, 0.00221]	912	0.00	0.00	0.00	0.50
2	(0.00221, 0.00459]	4707	0.00	0.01	0.00	0.14
3	(0.00459, 0.0126]	2808	0.01	0.01	0.00	0.20
4	(0.0126, 0.0254]	1127	0.03	0.02	0.00	0.19
5	(0.0254, 0.0427]	1274	0.04	0.03	0.02	0.61
6	(0.0427, 0.0787]	1935	0.07	0.07	0.01	0.08
7	(0.0787, 0.5]	891	0.12	0.14	0.03	0.19

Figura E.23: Resultado python calibración global. Fuente: elaboración propia

Calibración por país:

	país	n	mean_pred	mean_obs	abs_err
0	Argentina	31	0.17	0.06	0.10
1	Bolivia	18	0.01	0.06	0.05
2	Chile	14697	0.02	0.02	0.00
3	Colombia	607	0.04	0.05	0.01
4	Costa Rica	9	0.09	0.11	0.02
5	Ecuador	143	0.03	0.02	0.01
6	El Salvador	2	0.05	0.00	0.05
7	España	10	0.10	0.00	0.10
8	Guatemala	6	0.08	0.00	0.08
9	México	926	0.03	0.03	0.00
10	Nicaragua	6	0.00	0.00	0.00
11	Paraguay	6	0.01	0.00	0.01
12	Perú	85	0.01	0.00	0.01
13	United States of America	12	0.02	0.00	0.02
14	Uruguay	38	0.00	0.00	0.00
15	Venezuela	63	0.03	0.02	0.01

Figura E.24: Resultado python calibración por país. Fuente: elaboración propia

Calibración por plan:					
	plan_nombre	n	mean_pred	mean_obs	abs_err
0	Básico	7753	0.02	0.02	0.00
1	Enterprise	450	0.01	0.01	0.01
2	Individual	3128	0.04	0.05	0.01
3	Reservo	5298	0.02	0.02	0.00

Figura E.25: Resultado python calibración por plan. Fuente: elaboración propia

Calibración por rubro:					
	rubro	n	mean_pred	mean_obs	abs_err
0	Centro de estética	1966	0.02	0.02	0.00
1	Medicina / Salud	11031	0.02	0.02	0.00
2	Odontología	897	0.02	0.02	0.00
3	Psicología / Psiquiatría	1374	0.04	0.05	0.00
4	Salón de belleza / Barbería	522	0.04	0.04	0.00

Figura E.26: Resultado python calibración por rubro. Fuente: elaboración propia

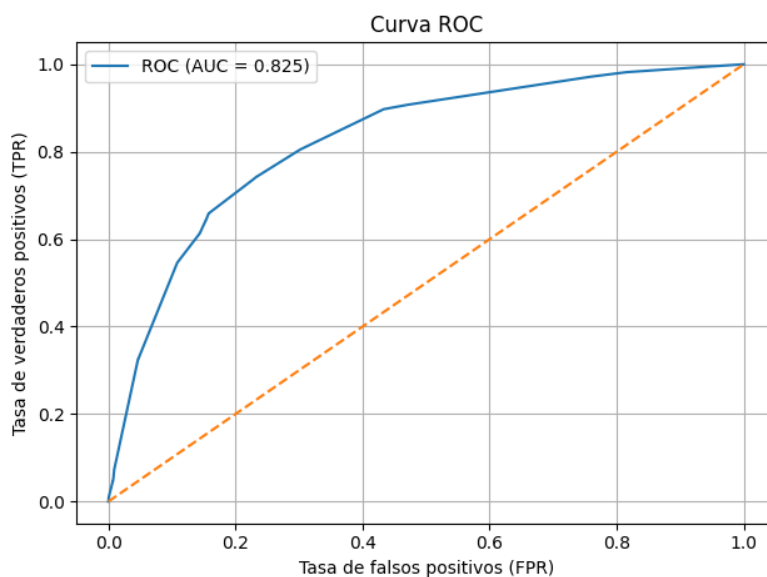


Figura E.27: Gráfico de curva ROC. Fuente: elaboración propia

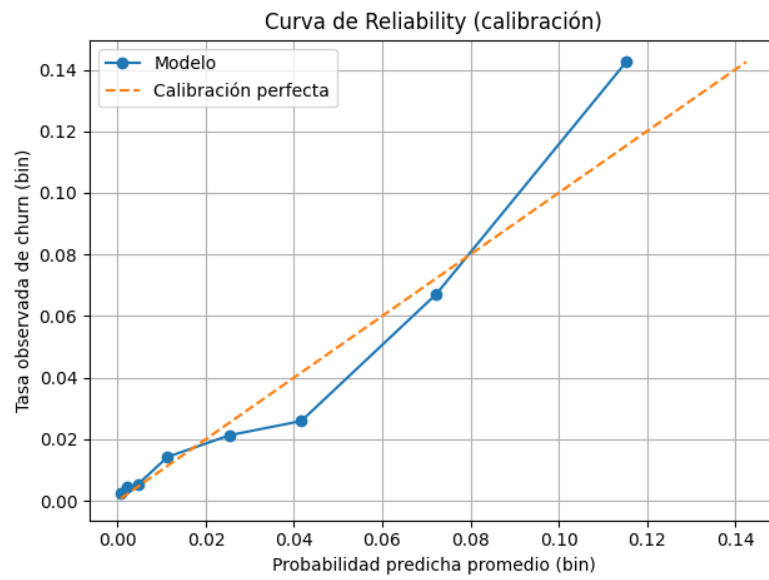


Figura E.28: Gráfico curva de calibración. Fuente: elaboración propia

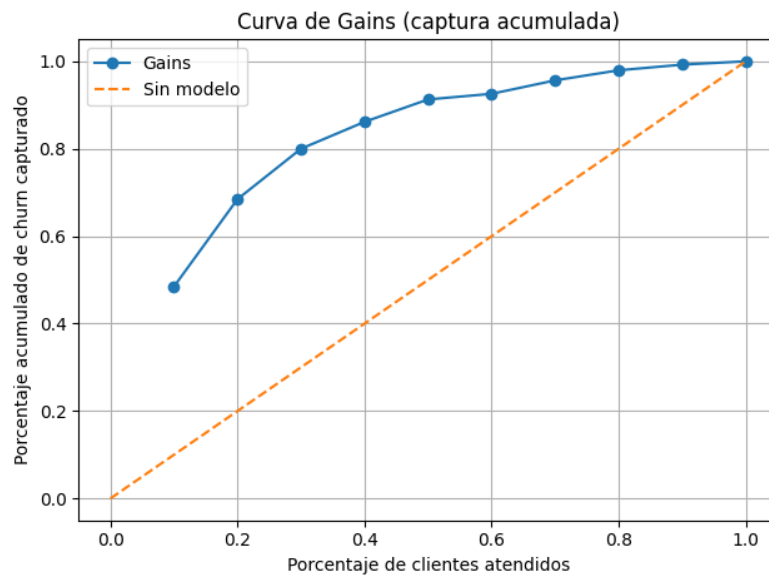


Figura E.29: Gráfico curva de captura acumulada. Fuente: elaboración propia

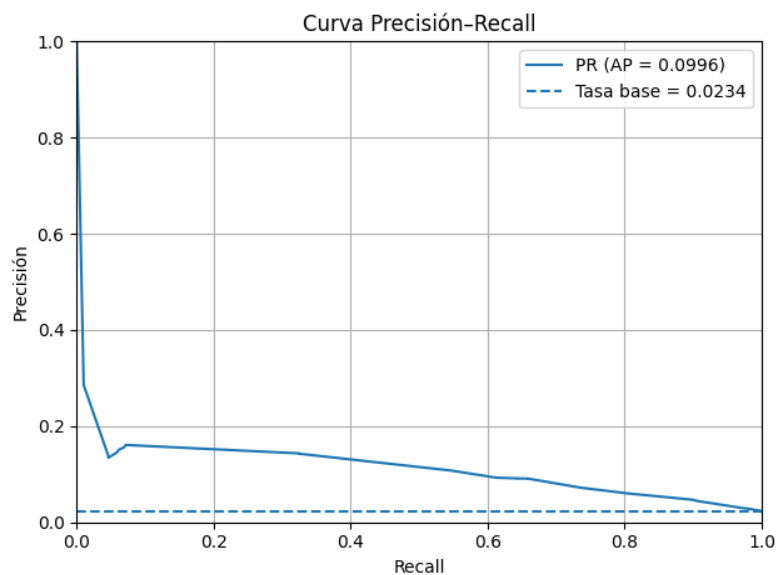


Figura E.30: Gráfico curva precisión-recall. Fuente: elaboración propia

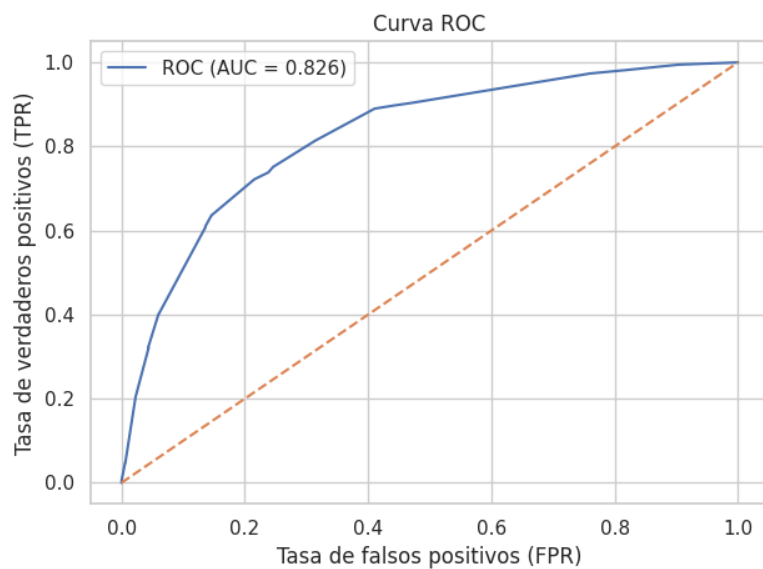


Figura E.31: Gráfico curva ROC Chile-México-Colombia. Fuente: elaboración propia

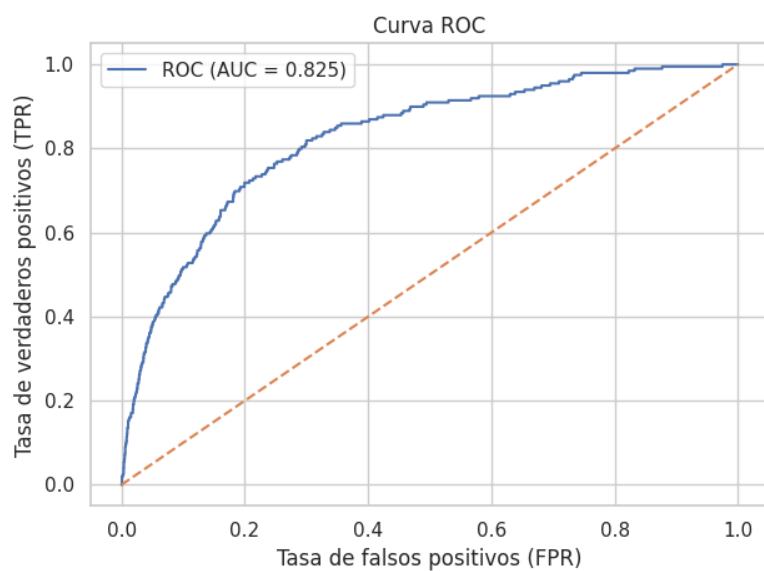


Figura E.32: Gráfico curva ROC Chile + Medicina. Fuente: elaboración propia

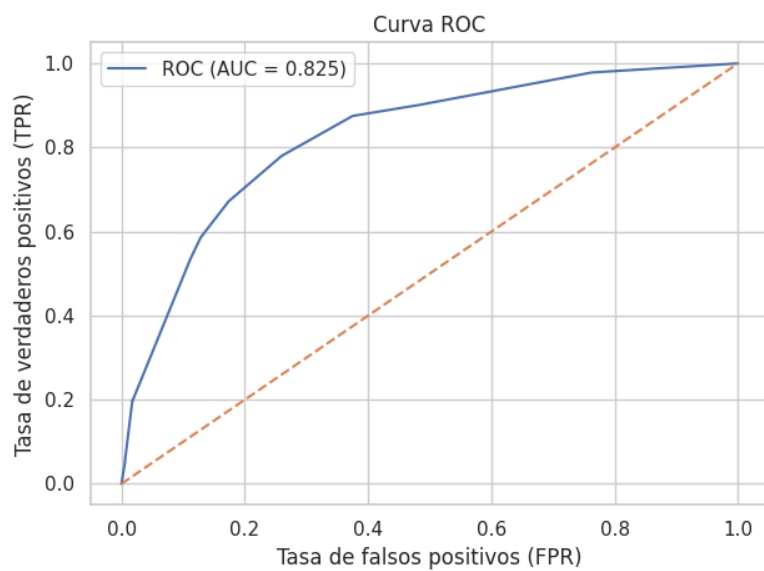


Figura E.33: Gráfico curva ROC medicina. Fuente: elaboración propia

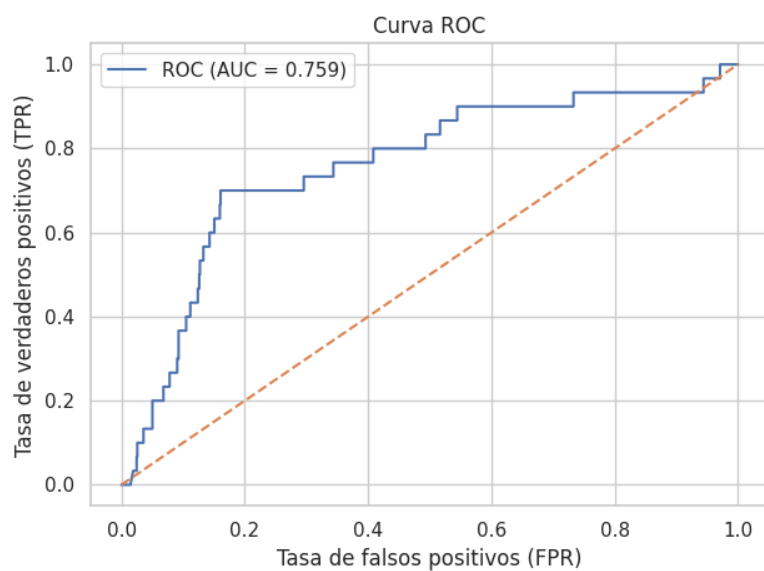


Figura E.34: Gráfico curva ROC México. Fuente: elaboración propia

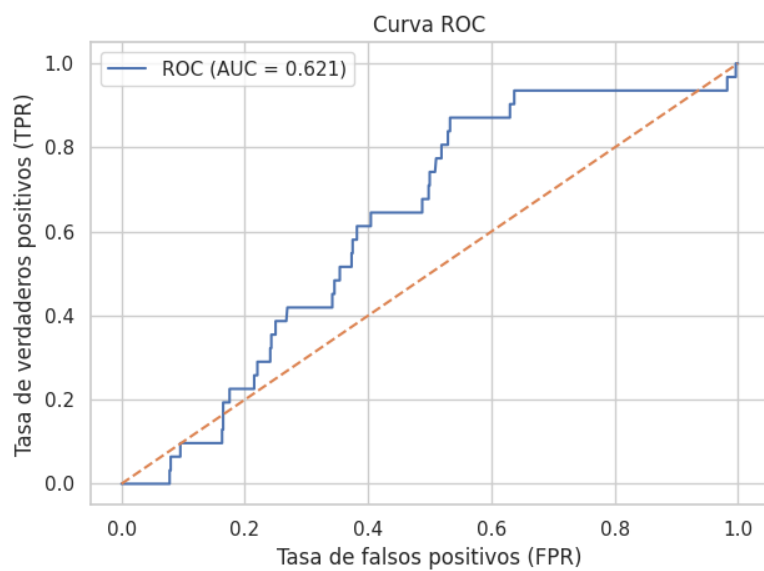


Figura E.35: Gráfico curva ROC Colombia. Fuente: elaboración propia

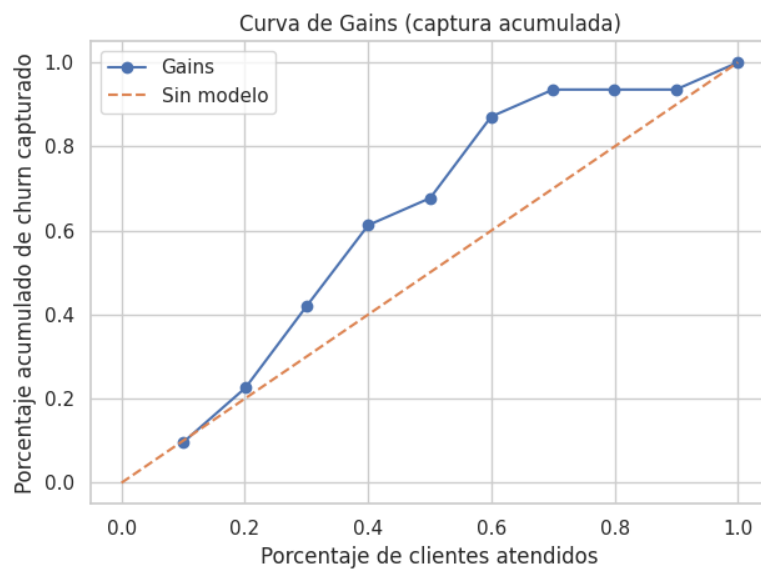


Figura E.36: Gráfico curva captura acumulada Colombia. Fuente: elaboración propia

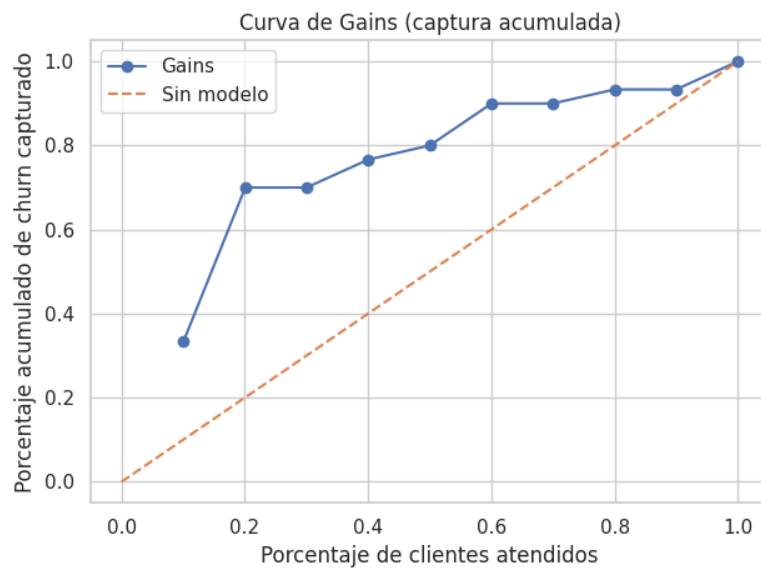


Figura E.37: Gráfico curva captura acumulada México. Fuente: elaboración propia

## Anexo F. Implementación

```

Colab para Customer Success ☆ ☁
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

mandos + Código + Texto ▶ Ejecutar todas

[ ] ▶ # === Celda 0: Utils de split temporal + métricas ===
import pandas as pd
import numpy as np
from sklearn.metrics import average_precision_score, precision_recall_curve
from sklearn.linear_model import LogisticRegression

# ----- helpers -----
_DEFAULT_FEATURES = (
    "arpa_proxy",
    "citas_30d",
    "recencia_uso_dias",
    "bloques_90d",
    "tenure_meses",
    "plan_nombre",
    "pais",
    "rubro",
)

def _check_features_exist(df: pd.DataFrame, cols):
    missing = [c for c in cols if c not in df.columns]
    assert not missing, f"Faltan columnas en df: {missing}"

```

Figura F.1: Colab de uso Customer Success. Fuente: elaboración propia

id_cliente	mes_corte	plan_nombre	pais	rubro	tenure_meses	arpa_proxy	prob_churn_T1	prob_churn_3m	riesgo_T1	riesgo_3m	pct_cartera
7473	01-11-2025	Individual	Argentina	Salón de belleza / Barbería	2	30	50%	88%	Muy alto	Muy alto	0%
7468	01-11-2025	Básico	Argentina	Psicología / Psiquiatría	2	60	50%	88%	Muy alto	Muy alto	0%
7604	01-11-2025	Individual	Argentina	Salón de belleza / Barbería	1	40000	50%	88%	Muy alto	Muy alto	0%
7859	01-11-2025	Básico	Argentina	Medicina / Salud	0	86000	16%	40%	Alto	Alto	0%
7844	01-11-2025	Reservo	Argentina	Centro de estética	0	90000	16%	40%	Alto	Alto	0%
7860	01-11-2025	Reservo	Argentina	Centro de estética	0	90000	16%	40%	Alto	Alto	0%
7868	01-11-2025	Individual	Argentina	Psicología / Psiquiatría	0	40000	16%	40%	Alto	Alto	0%
7828	01-11-2025	Básico	Argentina	Medicina / Salud	0	91000	16%	40%	Alto	Alto	0%
7851	01-11-2025	Básico	Argentina	Centro de estética	0	81000	16%	40%	Alto	Alto	0%
7502	01-11-2025	Básico	Chile	Medicina / Salud	2	40000	16%	40%	Alto	Alto	0%
7485	01-11-2025	Básico	Chile	Medicina / Salud	2	130000	16%	40%	Alto	Alto	0%
7487	01-11-2025	Reservo	Argentina	Medicina / Salud	2	70	16%	40%	Alto	Alto	0%
7520	01-11-2025	Individual	Argentina	Psicología / Psiquiatría	2	40000	16%	40%	Alto	Alto	0%
7492	01-11-2025	Individual	Argentina	Psicología / Psiquiatría	2	30	16%	40%	Alto	Alto	0%
7366	01-11-2025	Básico	Chile	Medicina / Salud	3	45000	16%	40%	Alto	Alto	0%
7423	01-11-2025	Reservo	Colombia	Psicología / Psiquiatría	3	210000	16%	40%	Alto	Alto	1%
7533	01-11-2025	Individual	Argentina	Medicina / Salud	2	30	16%	40%	Alto	Alto	1%
7398	01-11-2025	Básico	Argentina	Centro de estética	3	50	16%	40%	Alto	Alto	1%
7569	01-11-2025	Individual	Argentina	Centro de estética	2	40000	16%	40%	Alto	Alto	1%
7276	01-11-2025	Reservo	México	Medicina / Salud	4	5110	16%	40%	Alto	Alto	1%

Figura F.2: Excel de salida del modelo. Fuente: elaboración propia



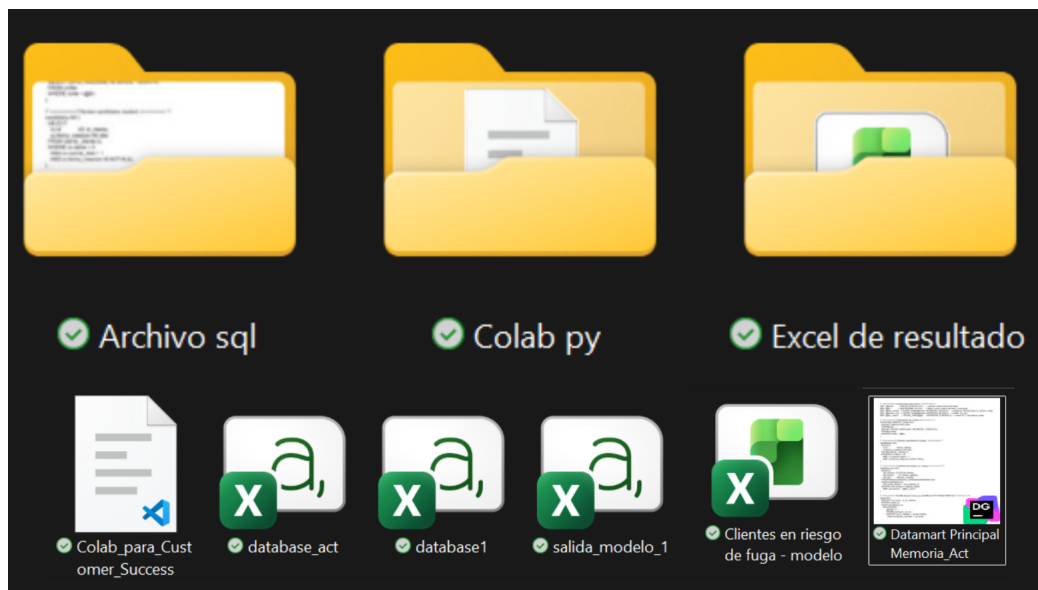


Figura F.3: Carpetas entregadas a Customer Success. Fuente: elaboración propia

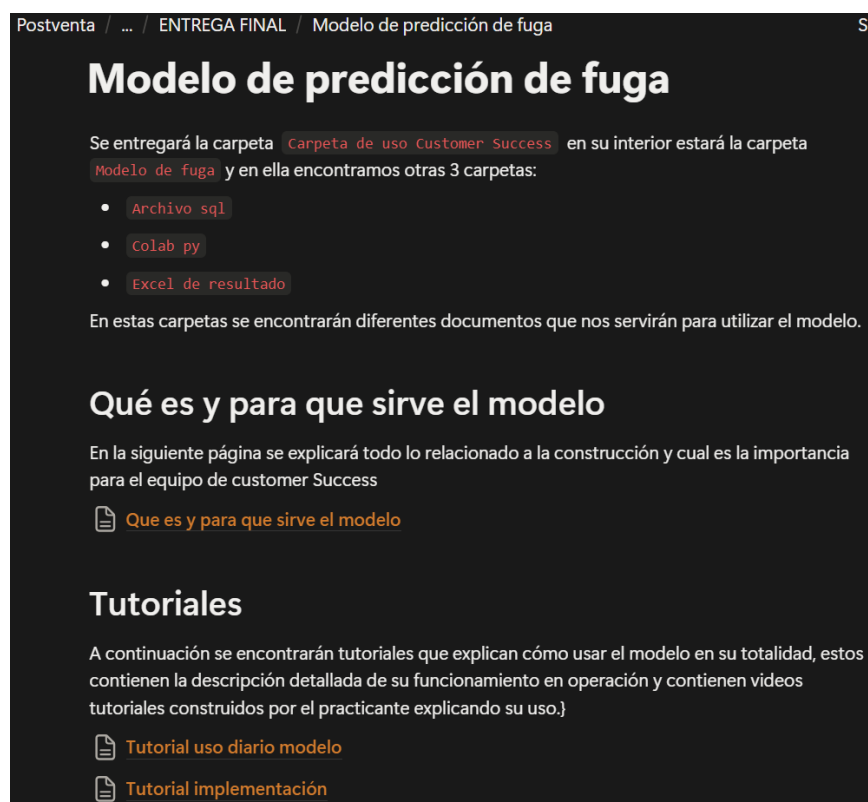


Figura F.4: Página de Notion elaborada para Customer Success. Fuente: elaboración propia

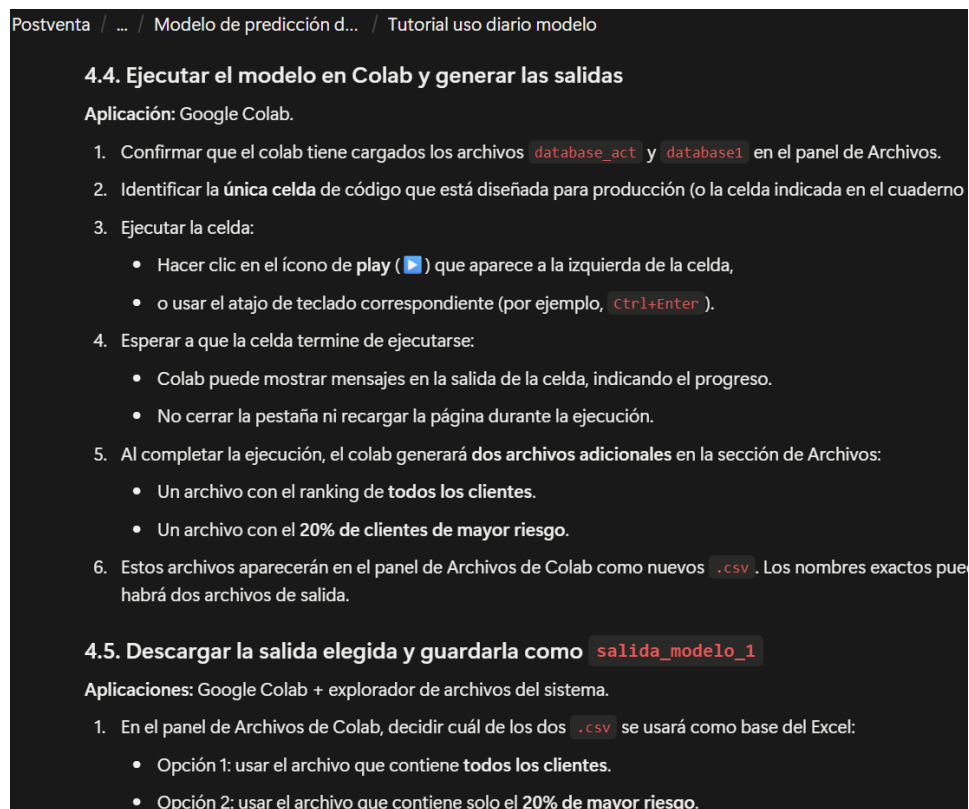


Figura F.5: Extracto guía de implementación y uso elaborada para Customer Success. Fuente: elaboración propia

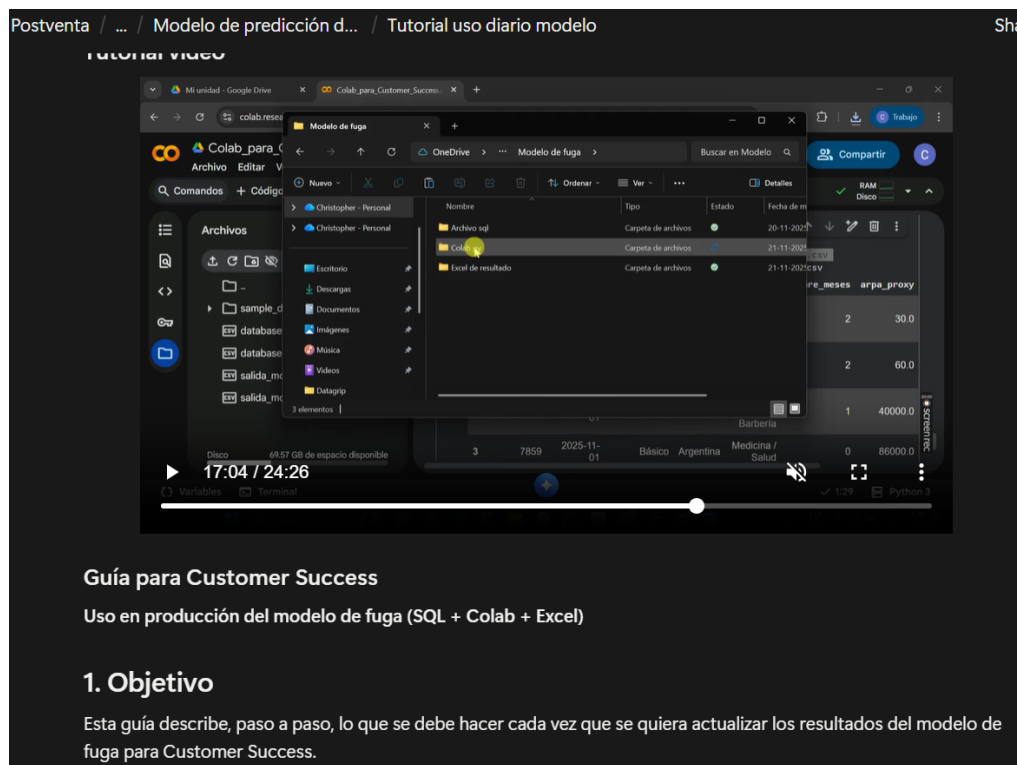


Figura F.6: Tutorial en video elaborado para Customer Success. Fuente: elaboración propia